JOURNAL OF PROJECTIVE TECHNIQUES

VOL. 21 SEPTEMBER, 1957 No. 3

• • •

EDITOR

BRUNO KLOPFER

EXECUTIVE EDITOR

BERTRAM R. FORER

ASSOCIATE EXECUTIVE EDITORS

WALTHER JOEL
EDWIN S. SHNEIDMAN

STAFF EDITORS

EVELYN CRUMPTON
FLORENCE DIAMOND
JAMES HOLT
NICHOLAS ROSE
HARVEY ROSS
LEONARD B. OLINGER
JEANNE M. REITZELL
JOHN WEIR
IRLA LEE ZIMMERMAN

CONSULTING EDITORS

JOHN E. BELL
EUGENIA HANFMANN
ROBERT R. HOLT
LOIS B. MURPHY
MARIA A. RICKERS-OVSIANKINA
L. JOSEPH STONE

The Clinical and Scientific Methods: Synthesis or Antithesis?[1]

JULES D. HOLZBERG
Connecticut State Hospital

After close to two decades of functioning as a clinical psychologist, a period that should be sufficient to have brought one through the development of one's professional self-concept, I must confess that I find myself still puzzled by the question, "What am I?", realizing only too well that confessions may be good for the soul but bad for one's reputation. To some the clinician is an artist, an artist who hides behind a scientific facade, who is blind to scientifically relevant facts, who responds with faith rather than with evidence. Still others see him as a scientist of the highest creative order, weaving individual personality theories, which may be at the rate of one to five per week. As if this has not been enough to provide confusion, it has now even been suggested that the clinician behave like a "cook" armed with a cook book (18), which seems to be no solution either, since has it not been said, "Cookery has become an art, a noble science?" The clinical method, like the clinician, has also been subjected to a battery of adjectival descriptions which are no less divergent:

> "The clinical method ... is labeled by its proponents as dynamic, global, meaningful, holistic, subtle, sympathetic, configural, patterned, organized, rich, deep, genuine, sensitive, sophisticated, real, living, concrete, natural, true to life, and understanding. The critics of the clinical method are likely to view it as mystical, transcendent, metaphysical, supermundane, vague, hazy, subjective, unscientific, unreliable, crude, private, unverifiable, qualitative, primitive, prescientific, sloppy, uncontrolled, careless, verbalistic, intuitive, and muddle headed." (17, p. 4-5).

This controversy has led me to consider the question, "What is the clinician?" and I have sought for the answer in what it is that the clinician does—the process of clinical inquiry. My focus has been on the clinician as a psychodiagnostician rather than as a therapist or as a researcher for the confusion I have described emerges from the contemporary controversy concerning what the clinician is as a diagnostician. Since projective techniques constitute the heart of contemporary clinical diagnostic practice, my search for clarity of professional self-concept seems relevant to those of us who have joined together in a common interest in projective methods. I am limiting myself to clinical diagnosis in the psychiatric setting which has been my primary experience and represents the contemporary accepted pattern of multi-disciplinary collaboration in clinical diagnosis. I have sought clarification to the question in an introspective examination of my own performance as a clinical diagnostician and in my observations of psychology interns whom I have had an opportunity to observe over the years. I feel that interns can especially contribute much to our understanding of this problem for it is in the embryo psychologist that one can observe most clearly aspects of the clinical process that may be more difficult to unravel in the more mature professional clinician with his highly developed professional defenses.

The process of clinical inquiry has often been thought of as a specific method of problem solving that could describe all clinicians' behavior, concerning the nature of which there would be general agreement. In actuality, is there such a thing as *the* clinical method, or are there a variety of clinical methods many of which differ only in the degree to which certain aspects are emphasized, yet by this

[1] Presidential address delivered at annual meeting of Society for Projective Techniques, September 1, 1956, Chicago, Ill.

EDUCATION LIBRARY

very difference in emphasis produce different methods in much the same way that we speak of unique differences in personality based on quantitative differences in personality variables?

As I view it, clinical inquiry consists of a series of interrelated activities which do not follow a prescribed sequence but rather overlap considerably. Recognizing the risk in so doing, I have nevertheless attempted to dissect the process of clinical inquiry in the belief that this will facilitate the solution to the problem of what the clinician is. I have isolated ten activities or variables which I believe to be pertinent to clinical inquiry.

The Variables of Clinical Inquiry

If one looks upon clinical performance as imbedded in a process of communication, it is the referral problem or question which initiates the communication process and leads ultimately to a communication, either a written or oral psychological report. The first activity relevant to understanding the process of clinical inquiry is the extent to which the clinician accepts the referral problem as the focal problem. Here, we may observe two sharply differing attitudes. One attitude is that of focusing on the referral problem, to the exclusion of other conceivable problems relative to the patient. The second attitude accepts the referral problem as but one of a series of questions that may ultimately be pertinent to the patient. Here the emphasis is on broad personality study that will ultimately lead not only to the answer to the specific referral question but to others that may become relevant. I believe that differences in this factor of the clinician's attitude toward the referral question is basic, in that this attitude will affect the other variables in the clinical process.

In the type of professional setting that I have earlier described, the questions that are presented to the diagnostic clinician emerge primarily from other professional disciplines. For the most part, these questions are posed in terms of those variables pertinent to the professional discipline involved, e.g., psychiatry or social work. This leads to the second factor in clinical inquiry—the reformulation of the referral question into the variables of psychology. Here again we may observe differences in the way the same question may be reformulated. These differences will be a function of the focal or non-focal attitude I have just described and of varied theoretical orientations and the differences in significance of variables which these orientations presume.

A third activity in the process of clinical inquiry is the selection of techniques which are capable of illuminating the psychological variables with which the clinician is concerned. Here again one may note differences in the types of techniques used — from techniques of specific variables to those of multiple variables, from techniques that yield scores to those that yield not-readily quantifiable data. There may be a single technique or a battery of techniques, standard batteries or flexible batteries.

The unique character of the clinical process is that it involves an interaction between two individuals, and it is this interactional process that I consider a fourth factor in clinical inquiry. Here one may note again differences in attitude toward the importance of this factor in determining the nature of data elicited and differences in the way knowledge of this interactional process is used in clinical understanding. Attitudes toward the importance of this factor range from that of complete indifference to it to insistance that no data can be analyzed validly outside such a frame of reference. Differences in the use of knowledge emerging from the effects of the interactional process are principally due to lack of understanding of how different relationships affect clinical data and what

these effects mean in terms of patient understanding.

From the point at which contact with the patient begins, the clinician has begun his process of data collection, the fifth variable in the process of clinical inquiry. There are multitudes of data in the clinical situation but each clinician operates within a perceptual frame of reference that permits selection of certain observations and not others, depending on his theoretical orientation, the psychological variables relevant to the formulation of the problem and the clinician's own defensive orientation to reality. Here again are differences in another facet of the clinical process.

Both clinician and non-clinician agree, whatever else their differences, that the clinician is a vital variable in the clinical process. As Schafer has said, "No matter how helpful a clinical tool it may be, a psychological test cannot do its own thinking" (27, p. xi). But even prior to the process of interpretation, while the clinical tool may elicit observations, only the examiner can select them in terms of pertinence to the goal of the investigation. I have therefore considered him the sixth variable in the process of clinical inquiry. No special emphasis is needed to stress the differences in clinical performance resulting from differences in clinicians—differences in receptivity, integrative capacity, flexibility, capacity to benefit from experience, to name but a few.

The heart of the process of clinical inquiry is interpretation, the seventh factor. Here again one may note differences which may affect the final outcome of clinical inquiry. One may observe differences in modes of interpretation and levels of interpretation. Variations in modes may be noted by interpretive emphasis on authority, on subjective norms and/or psychological theory. Variations in levels are noted in interpretations limited to isolated behavior, interpretations extending to the formulations of discrete hypotheses and interpretations involving coordinated hypotheses resembling in many respects a specific theory. Even in the process of hypothesis construction, one can note hypotheses evolved through a systematic inductive process, hypotheses evolved as tentative hypotheses and tested deductively with new data, and hypotheses evolved seemingly as a result of a sudden insight. Perhaps in no other facet of the process of clinical inquiry can differences be so readily detected.

The eighth variable in the process of clinical inquiry is the making of predictions. Here again are variations, of which there are at least two general types which may be clearly delineated. One of these is where the prediction is made directly from the data without an intervening process of interpretation. This most resembles the clinical actuarial or "sign" approach. Here the data are used as direct sign-posts to prediction. The other mode of prediction proceeds from the data to formulated hypotheses and then through a series of logical deductions. The predictions here are the end-products of the time-honored hypothetico-deductive process.

The ninth variable in the process of clinical inquiry is the determination of the success of predictions. This is one of the most neglected phases of the process of clinical inquiry and here too there exists wide differences in the extent to which such evaluation is conducted as part of the process of clinical inquiry and in terms of the criteria used to judge success or failure of prediction.

We began our investigation into the nature of clinical inquiry by stressing that it is embedded in a communication process which begins with the question posed by a non-psychologist and finally reaches the stage of the communication itself, i.e., the written or oral psychological report. It is this communication that constitutes the tenth variable in clinical inquiry. One need not stress the variability that exists in this process of communication, e.g., variability in terms

of clarity, comprehensiveness and usefulness.

I have briefly enumerated ten factors that I believe are variables determining the nature and quality of clinical inquiry. Perhaps I have belabored this point too extensively but I have wished to give substance to my conviction that clinical inquiry is not a unitary process except in terms of goal, i.e., understanding of the patient. Rather, there are many methods of clinical inquiry, some of which are startlingly dissimilar in their basic philosophy and conceivably in the results that they achieve.

The Nature of Scientific Inquiry

Does this mean that it is not possible to deal with the issue of what is the clinician? Artist, scientist, cook, what have you? I believe it complicates the problem but does not render it impossible. We might begin by asking whether the facets of the clinical process that we have dissected are identical or similar to those of scientific inquiry. Actually, as you will note, they are remarkably alike. Scientific inquiry is problem-solving activity that begins with a question often naively conceived by a non-scientist which requires reformulation into the variables of the scientist performing the research. Once so reformulated, relevant techniques for the elicitation of the variables are selected and data collection is begun. Interpretations in the form of hypotheses are advanced, sometimes at the beginning of scientific inquiry, sometimes at the conclusion, sometimes during the process of scientific inquiry itself. The process of interaction between clinician and patient does not loom as significantly in, say, experimental psychology. But this may be due to the failure of experimental psychology to recognize the importance of the variable of experimenter-subject interaction. If experimental psychologists working with humans do not stress this, it is at least encouraging to note that animal psychologists are becoming concerned with the nature of the relationships between animal and experimenter (23). Similarly, there are requisite skills and attitudes required of scientists as there are of the clinician. And, of course, scientific inquiry operates with predictions and tests their validity. Finally, scientific inquiry concludes with a communication that shares its findings with others.

Shall we therefore conclude from this that clinical inquiry and scientific inquiry are essentially identical? Some of us have reached that conclusion but I believe they are like Humpty Dumpty in *Alice in Wonderland* who says, "When I use a word it means just what I choose it to mean—neither more nor less." Scientific method is a process of problem solving with built-in provisions for maximizing validity. The success of the method has quite validly earned for it high prestige value, and it is therefore understandable that some will wish to identify with it. This however does not establish the validity of the identification.

Before proceeding with this issue, I would like to introduce to you an old friend of mine, Joe by name, who runs an auto mechanics shop in my home town. Joe advertises that he utilizes scientific methods of diagnosing the anatomical and physiological ills of motor cars. I have spoken to Joe and he is quite sincere in his belief that he is a scientist in his methods of diagnostic appraisal. I called on Joe recently to tell him my car would not go. I proceeded to describe the symptoms of my car's ills with my naive understanding of auto mechanics which Joe proceeded to translate into the variables of the car mechanic that led to three alternative hypotheses—something was wrong with my car's ignition system, the battery, or the fuel pump. Instruments were then applied to my car—instruments which I cannot describe or understand but which nevertheless were his techniques

of study and yielded both quantitative and qualitative data. Through his techniques and his careful observations of my car, his process of data collection began. It was clear that he had certain skills and personality traits that made for success in his diagnostic role so that he was a necessary part of the process of his inquiry. Inherent in his hypotheses were the predictions that he made, each prediction serving to support one of the hypotheses. Where predictions relative to one hypothesis were not supported by his study of the car, this hypothesis was discarded and at long last one of the hypotheses was supported and treatment could be begun. Lastly, he shared his findings with his associates and even recorded his diagnosis for future reference. He subsequently communicated his charges for his services, but I would suggest that this is not vital to our concern with clinical or scientific inquiry. Is Joe a scientist? He seems to behave like a scientist but I doubt that we would agree that this is his correct occupational designation.

The picture of what Joe is becomes clearer when I report that on the following day my car developed the symptom of "coughing." I observed Joe again as he proceeded to diagnose my car and to describe his behavior would be to repeat his performance of the day before. Here it became apparent that Joe's confirmed hypothesis of the day before was no longer relevant to my car's ills today. Unlike the true scientist who is involved in evolving hypotheses and eventually integrating them into a frame of theory, Joe was concerned with a narrow explanation of a specific problem and not in evolving a comprehensive theory of my car's personality. The clue therefore as to whether Joe the mechanic, Sam the tailor, or Harry the clinician is a scientist or not lies, it seems to me, in understanding their roles as theory builders in relation to their subject matter areas.

STRATEGIC HYPOTHESES VS. TACTICAL EXPLANATIONS

What is the nature of the hypotheses developed by scientists as compared to non-scientists? The hypotheses of the scientist are directed toward the ultimate development of comprehensive theories. They are open-ended in that they are capable of change and growth. They should be able not only to answer specific problems but to lead to new knowledge and understanding. They should be increasingly capable of inter-connecting facts which were otherwise disconnected. They should constitute logical next steps in the systematic development of a theory. *These are "strategic" hypotheses rather than "tactical" explanations* (20). They are not merely concerned with a specific question but a higher ordering of facts.

In this context that stresses the pertinence of the nature of hypotheses to science, can the clinician function as a scientist in his role as psychodiagnistician? Obviously he can, when his concern is the development of *strategic hypotheses* that will ultimately lead to a comprehensive theory of the individual he studies. Do all clinicians function as scientists? I believe not. Many, if not most clinicians, are not concerned with the development of *strategic hypotheses* about a patient but view their goal as the handling of a specific, more narrowly-defined problem. This lack of concern may be based on ideological differences with regard to the centrality of hypothesis–construction to science but more importantly I believe it is based on problems of time economy. Theory building is never simple, never swift. Theory building in clinical practice is costly. To truly develop *strategic hypotheses,* not only must there be time to collect data (White (34) has suggested a minimum of ten to fifteen hours in comprehensive personality study) but time to think, without which hypothesis-construction is impossible. I am

therefore stating that the clinician is capable of functioning as a scientist in his diagnostic role, but that such functioning may be more the exception than the rule. Perhaps it may be more correct to state that he strives to be the scientist he is capable of becoming but he is often unable to achieve this role because the practical nature of the clinical situation serves to restrict him.

I have attempted to stress that the nature of the hypotheses advanced determines a problem-solving method's identification with science. But, the development of *strategic hypotheses* is not science itself. It becomes part of the method of science when it is embedded in a process of inquiry that begins with a problem posed and proceeds to a communication. But I have also stressed that there is no single method of clinical inquiry. Can a model of clinical inquiry be offered which provides for the development of *strategic hypotheses* and also have built in the greatest assurance for validity of its final predictions, which after all is the ultimate goal of scientific inquiry? I believe such a model can be developed and I should like to tentatively suggest one by dealing with each of the factors of the process of clinical inquiry that I have earlier described.

A Model For Clinical Inquiry

The Referral Question

The ideal model for clinical inquiry begins with the question posed as but one of a series of questions that will ultimately be pertinent to the patient. Therefore, it emphasizes broad personality description that will lead to *strategic hypotheses*.

Reformulation of Problem

This model recognizes that problems presented to the psychological clinician are in terms of variables that may be foreign to psychology. Therefore, it requires that each problem be re-formulated in terms of psychological variables. The question put to the engineer as to whether a bridge can support a given load must be capable of translation into variables that are pertinent to his field, i.e., stress, tension, etc. I belabor this point because it has seemed to me that clinical inquiry that proceeds from the question posed by the non-psychologist without the intermediate step of reformulation into psychological variables introduces a serious type of error into clinical inquiry, i.e., conceptualizing problems in terms of variables that are not relevant to the variables of psychological theory and psychological techniques. Unless the psychologist can so transform the question that is posed, it is not a problem appropriate to his function. We must learn "it is not every question that deserves an answer," at least if the question cannot be transformed into our own variables.

It is to be expected that there will be differences between clinicians as to their reformulations of questions depending upon differences in theoretical orientations and the differences in significance of variables which these orientations presume. The problem posed, "Is the patient schizophrenic?", may be reformulated by one psychologist into "Is this a patient who shows evidence of thinking disorder?", where the theoretical orientation is principally toward a descriptive psychopathology. The same question may be reformulated by another psychologist to, "Does this patient show indications of severe ego regression?", where the psychologist conceives of schizophrenia within a framework of a psychosexual dynamic orientation. Questions that are reformulated in terms of the broadest use of practical knowledge and theory provides the most adequate beginning of clinical inquiry. As in all forms of inquiry, the question cannot be transformed by anyone except the individual who is to perform the inquiry.

I think of a question often asked of the psychologist in an institutional setting, "Is the patient ready for dis-

charge?" This question has no significance for the psychologist unless it can be transformed into a problem whose variables are meaningful to him. Thus, for the patient who is admitted because of a suicidal attempt, the question might be reformulated as, "Are the patient's self-destructive attitudes altered by his treatment or are his controls more effective in dealing with them?" For another patient who is hospitalized because, since the death of his mother, he has not had anyone to supervise his long-standing schizophrenia, such a question could be transformed into, "Are this man's ego processes such that he can be expected to function with reduced distortion of reality?" If the question, however, should become transformed into the problem, "Since this patient is unchanged in his psychopathology, can he function in another environmental setting similar to the one that was available to him when his mother was alive?", then such a question seems more appropriate for evaluation by the psychiatric social worker and would therefore not be an appropriate question to be presented to the clinical psychologist.

The problem, now reformulated, becomes the goal of the inquiry but it must be stressed that this goal may be altered by the course of the inquiry itself which may indicate the irrelevance of the original question or may demonstrate more crucial problems that need investigation. Even where the original problem may remain throughout the course of the inquiry the goal toward which the inquiry is directed, once this problem has been answered new problems may emerge for which the psychological clinician is asked for answers. The process of clinical inquiry begins with a question that is reformulated in psychological variables and is constantly concerned with a question throughout the course of the inquiry. It should be stressed that the capacity to raise questions during the course of inquiry is not one that is possessed to the same degree by all clinicians. It has been said, "It is the work of scientific genius to be sensitive to difficulties where less gifted people pass by untroubled by doubt" (3). Clinical inquiry at its highest level requires this capacity to formulate and constantly reformulate the question or questions that guide clinical inquiry.

Selection of Techniques

The ideal model selects techniques that are capable of illuminating the psychological variables with which the study is concerned. Of single greatest importance here is the use of batteries of techniques, which perhaps is the most significant development recently in the practice of personality study (34). This model views the use of batteries as one of the great strengths in clinical inquiry by creating circumstances under which samples of behavior can be checked against each other. It stresses the use of techniques which permit the observation of interacting variables and denies that tests of single psychological variables that attempt to study these variables in isolation is either clinically useful or theoretically possible. This model stresses the use of techniques that yield scores as well as verbal and behavioral content and denies that a lack of primary interest in scores as such makes for a less objective method of study unless one uses the term "objective" in a pedantic sense rather than in the sense of primary concern with the "object" of study—the patient rather than the score (12).

One of the critical problems in clinical practice is the extent to which there should be standard batteries, and to what extent should batteries be determined by the reformulation of the problem. Thus, Rapaport (21) suggests a fairly intensive routine battery and offers a reasonable rationale for clinical practice being organized in this manner. This raises the serious question whether in so doing our testing techniques are becoming the center of the clinical process rather than

the problem determining the course of clinical inquiry. To me this does not present as serious a problem as it does to some if batteries such as the one recommended by Rapaport are the basis for our discussion. These batteries are recommended on the basis that they provide the opportunity to observe a multiplicity of interacting variables. If the battery is selected from the point of view of providing adequate samples of behavior of a patient, among which are presumed to be the particular variables with which we may be concerned, this structures the appropriateness of the use of the test battery. The presumption in the use of such a standard battery is that we are dealing with major dimensions of human personality which should be relevant for the study of all problems.

There is the danger in any area of inquiry that one may become too technique-oriented. It is quite easy to become enamored of instruments of measurement or observation but it is obvious that such instruments do not make for a field becoming scientific or disciplined—they are merely *means* to an end. Maslow (14) recently has indicated that American psychology has tended to become too technique-oriented. He points to some psychologists who will pursue a problem if it can be studied with animals or the psychologist who will perform a study providing he can subject the data to factor analysis. Here one sees techniques holding priority over the nature of the problem, which is contrary to true creative scientific inquiry. Our techniques are indeed critical but we must never lose sight of the fact that they are still techniques which cannot be worshiped for their own sake but only because they permit solutions to problems at hand.

Interactional Process

This model views the interpersonal relationship between examiner and subject as a requisite for clinical inquiry (5, 8, 19, 25, 26). It challenges the emerging stress by some that patients can be studied as effectively in group testing situations as in individual testing situations. This idea will probably never go out of style; it will be ridiculous year after year. *Clinical* inquiry, if it means nothing else, involves a contact between two people. To consider clinical inquiry on any other grounds is to this model a perversion of the meaning of "clinical." Now, it is true that until quite recently, this interaction was considered insignificant in clinical diagnosis except as it determined the representativeness of a subject's performance on a test by consideration of the subject's conscious attitudes toward the testing experience. This was essentially the meaning of rapport. But with this model, the psychological diagnostic examination is viewed as an interactional process, not only between subject and task but also between subject and examiner and examiner and subject. It is accepted that personalities are "open systems" which are in constant interaction with the environment which includes not only the examiner, but the nature of the situation, the implications of being tested, etc. It is White's (34) suggestion that perhaps batteries of examiners may here be as important as batteries of tests in order to more fully sample the nature of a patient's interactional behavior.

Deutsch (4), approaching the problem from the point of view of a field theorist, sees the test situation as one in which the clearly defined roles of examiner and subject involve certain basic social prerequisites and consequently test interpretation cannot be adequate unless it is within the framework of the social nature of the test situation. He states, "The projective test situation is a social situation in which the responses or behavior of the subject reflect his perception or subjective definition of the situation, his intentions in the situation, and his tendencies to behave in situations which are subjectively defined in the

way that he has subjectively defined the test situation. The subjective perception of the situation will be determined not only by his personality but also by such situational factors as the following: the characteristics of the task with which he is confronted, the interpersonal setting created by the personality and implicit expectations of the examiner, the social statuses of the examiner and the subject, the institutional setting and function of the test, and the cultural definitions of appropriate behavior in the testing situation" (4, p. 433-4).

Schafer (27) approaches this problem from the point of view of psychoanalytic theory and practice and attempts to relate the phenomena of transference and counter-transference to the clinical test situation. He has portrayed the interpersonal dynamics in the test situation in terms of the needs and problems of the tester as well as the patients. The import of these recent discussions is not only that there is an interpersonal interaction in diagnostic testing but the nature of this relationship has significant bearing on the data procured and therefore on the interpretation that the psychologist makes of his data. I know that there are some who deny that this relationship bears any significant role in determining the kind of data that will be elicited, while at the same time these do not deny that the clinical method involves an interpersonal process. A fair number of good research investigations seem to indicate, however, that this interpersonal relationship may not have a trivial effect on a subject's performance (1, 9, 11, 13, 24, 29, 32).

If this be so, does this lead to the conclusion that clinical techniques are without worth because of their presumed "unreliability" since varying the examiner should theoretically introduce variations in the subject's responses? This of course has been one of the sharp barbs most directed at projective techniques because their very nature makes them more sensitive to the interactional process. I would venture to predict that any test, clinical or non-clinical, is subject to the effect of this interpersonal equation although to varying degrees, reflecting the sensitivity of the clinical instrument to the subtleties of personality and the test situation. McClelland (16) has recently argued that the need for "test reliability has been greatly exaggerated by American psychologists."

Clinicians do not look upon this interpersonal equation and its effect on performance as a weakness of the clinical method but rather a strength that enriches the understanding of the subject. To be sure, we need to understand more the nature of this relationship and how it does effect performance; and we need to know how to interpret this for the purpose of patient understanding. The knowledge of the role of the interpersonal equation in clinical diagnosis can extend the range of data that is now available to the diagnostic clinician.

The knowledge of these interactional influences has changed the character of supervision in professional clinical practice. The emphasis has shifted from teaching principally test interpretive skills and now includes understanding the nature of the interactional processes and the way these may have determined the subject's responses. To be sure, much of the interactional phenomena are based on irrational ideas and attitudes in both patient and clinician which are beyond superficial comprehension but knowledge that such interaction plays a role can help avoid some of the predictive errors which clinicians make. The need is for more research to determine the effect of what interactional variables on what test behavior and on what subjects. The importance of these considerations have led to what I hope will some day become standard clinical practice—the interview following the completion of testing in order to clarify the nature of the interpersonal

equation that existed during testing.

Collection of Data

In this model, the emphasis is on the broadest collection of observations so that we do not draw "sufficient" conclusions from insufficient data. It stresses that ideally no observation essential to the psychological variables being studied is omitted. While this model does not confuse the myriad collection of unrelated data with science, it does recognize that clinical inquiry is initially a process of hypothesis formulation in which every datum is relevant. Once the hypotheses have been established, data collection is directed by the hypotheses themselves so as to establish their validity.

The Clinician

This model views the clinician as crucial to the process of clinical inquiry since it accepts the thesis that no psychological instrument can make the observations nor do the thinking that is required in the problem-solving task of clinical inquiry. It sees the clinician as one who is constantly questioning while searching for explanations in clinical study, who must be able to tie together diverse data and observations, who is capable of adapting himself and his techniques to the patient and the special problems he presents, who will benefit from his experience because he wants to learn from it, who can face contradictions as well as confirmations in his study of a patient. Not unlike the requirements of psychotherapy, he must be able to listen to his patient at various levels, to the usual meanings of his words, to the symbolic and latent implications of his verbalizations, to his behavior, his attitudes, his movements.

Interpretation

The heart of the process of clinical inquiry is the process of interpretation. Schneidman (28) has demonstrated that there are certain common steps in the clinical analysis of the TAT which are performed by different examiners. However, it has been my observation that there is no single interpretive process used by all clinicians. Rather, there are many such processes, some preferred more than others by certain individuals and in all probability all are used to some extent by most. If this be so, then the question of which is the best method of interpretation can only be answered by research where it would be possible to structure the necessary controls to insure appropriate evaluation. I shall here attempt to describe these interpretive processes as I see them and in so doing to caricature them since it is hardly true that any one method is used in isolation.

Before considering these types, we might consider what the object of the interpretive process is. Clearly it is to evolve an explanation that will answer the question raised at the beginning of the clinical inquiry. What is the nature of the explanation in the study of human behavior? It is in the form of an hypothesis that the clinician advances which meets the test of accounting most adequately for the observations that have accrued. It is within this context that the process of interpretation has been referred to as a process of hypothesis construction.

I have tended to see this process of hypothesis construction beginning with the referral, and the initial formulation of the problem implying a tentative hypothesis or hypotheses which can be tested by observations and data. The very first tentative hypothesis emerges from knowledge of what is already known about the patient, from the experiences and theories concerning the patient's problem advanced by others and by the previous experiences of the clinician himself. Where the problem is such that there is nothing of relevance known about the patient, where there exists no reported experiences of others or theoretical formulations, or where the clinician has had no prior experience with such a problem, a tentative hypothesis cannot be formulated unless

it be in the form of an intuitive hunch. Some clinicians will resort to this intuitive hunch to direct their inquiry. Others will avoid such an attempt at even tentative explanation and engage in a process not unlike an exploratory study. They will interview the patient briefly, use a test or series of tests that they have learned help to chart a course for their inquiry but in either case there comes some point in the clinical process where an initial tentative hypothesis or hypotheses begin to emerge and the further direction of clinical inquiry is to elicit observations and data that will disprove the initial hypothesis or modify it.

This model views the clinical process as a continuing process of hypothesis reformulation leading to the accrual of new pertinent evidence and insights. One significant role of test batteries is to provide new data pertinent to the hypothesis being considered. As in all forms of inquiry, one does not prove an hypothesis; one fails to disprove it and in clinical inquiry the hypothesis stands if it cannot be disproved. To some, it may appear that it is unscientific to develop an hypothesis before all the "facts" are in. In actuality, scientific research can begin with an hypothesis, conclude with an hypothesis, or develop one in the course of inquiry (10).

How do hypotheses emerge? What is the nature of the thought experiences that leads ultimately to that first hypothesis? Actually, little is known of this, the most creative aspect of the process of clinical inquiry. Some have even suggested that this cannot be subjected to study and to do so would be irrelevant. Their concern is with the validity of interpretation, not the means by which this is achieved. However, it is my contention that there is no single method of clinical interpretation, and until we understand the process more adequately, we cannot know which method is more valid. Furthermore, there is no reason to believe that the creative process which is involved here is any less subject to study than any other form of creative behavior. To the clinician who begins with an intuitive hunch, the presumption is that this cannot be explained on rational grounds, that this is an entirely subjective process of interpretation. Whether this is as truly subjective as some have stressed may be subject to some question. Cofer (2), in an analysis of clinical intuition, concluded that intuition means "unanalyzed, unrecorded, and perhaps unverbalized observations by the clinician." But where hypotheses emerge not as intuition but from data and observations, what is the nature of their emergence? Obviously the hypothesis is not in the data and observations. Rather the clinician does something to them which then emerges as a hypothesis.

This model recognizes two methods of hypothesis construction. One is essentially unplanned, a fumbling process in the search for explanation, finally emerging as does insight (7). The other method of hypothesis construction is a process of slow, orderly, inductive reasoning where data and observations are slowly integrated into an explanatory proposition which becomes the hypothesis. Tomkins (33) has indicated that the time-honored canons of inference can be applied to TAT data, thus demonstrating that the formal logic of inductive science can be applied to clinical data.

What guides the nature of the thinking process as the psychologist develops his hypotheses? This model recognizes three methods whereby data become meaningful. First, is interpretation by authority. This is the most prevalent in the lives of our younger colleagues but is represented in most of us. In using this approach, the clinician is saying: the authority has a vast body of experience in which I can share vicariously. These may be in the nature of formal objective norms or subjective norms which represent an important

accumulation and summarization of vast experience. The authority has also used his experience to develop theoretical conceptualizations of the meaning of data and this theoretical thinking is available for the interpretative process. In essence, this is not unlike the assumptions that are made in supervision—the supervisor like the authority has subjective norms and a better comprehension of theory relative to the tests, and the trainee shares vicariously in the supervisor's knowledge. The difference is that the supervisor is dealing with his patient while authority is dealing with classes of cases of which his case can only be an approximation. This represents the chief source of error—generalizations which may not embrace the particular patient. This method of interpretation has led Rapaport (22) to caricature those clinicians as operating "with the help of texts used as dreambooks" so that unbridled fantasy of the dream operates or each datum yields the same interpretation patient after patient. As it has been said, "books are good enough in their own way but they are a highly bloodless substitute for life." This model accepts sharing of experiences and ideas as of importance in interpretation, but emphasizes that the clinician deals with the facts as they are rather than what someone has said they ought to be.

The second method is interpretation by subjective personal norms. This is interpretation of data in accordance with one's personal experience. Here there is presumed to be an accumulation of experience which has resulted in some categorization, however simple, which serves as subjective norms for interpretation. The individual says, "I have seen this occur several times before in a certain category of patients and this helps me to understand this particular patient." This model accepts the importance of personal experience in clinical interpretation but recognizes that the norms of the clinician may be unreliable for a variety of reasons and until we understand more as to how an individual accumulates such norms and uses them we are not in a position to know how such norms may be improved. It may be that experience for some of us teaches much, but "learns us little."

The third method is interpretation by theory. This model recognizes that an important mode of clinical interpretation is through the intelligent and valid use of theory based on the presumption that data about personality can be understood by the application and use of personality theory, here defined in its broadest possible terms without reference to any specific theory. But this too has its dangers, not the least of which is the tendency of theory in some hands to become a straitjacket that stifles the creative process of clinical inquiry. This is to be seen most forcefully in the stereotypist who uses theory but to create stereotypes in the image of his theoretical premises. This is reductionism at its extreme.

This model views theory as a compass that charts a direction but not as a road map that tells you where you will finish the trip before you have decided where you are going. It does not reduce the clinician to the role of a technician who has learned a formula and this formula is the "Open Sesame" for all inquiry. It charts direction—it directs us away from the improbabilities to the probabilities but leaves the clinician with his most precious possession, the capacity for creative thinking. This model stresses that clinical interpretation involves creative thinking at its height and has no place for him who is afraid to think. Theory is general but the patient is specific and the specificity lies not in the theory but in the data which represents the uniqueness of the patient. The clinician does not ignore data because they are not stressed by theory. He does not hesitate to cut away ruthlessly those theoretical concepts that are inconsistent

with his data (12).

While this model accepts a number of modes of interpretation, it has only one criterion of when hypothesis building is complete and that is when the hypotheses have successfully accounted for all of the available data and observations. In other words, when the hypotheses are internally consistent in that the data can be subsumed under the structured hypotheses. The clinician must be aware that where other hypotheses may be equally valid, further inquiry is necessary. However, he is never absolutely certain of his hypotheses as in all scientific inquiry. The test lies in predictions that they permit and in the validity of these predictions.

This model, as we described earlier, views the development of *strategic hypotheses* as consistent with scientific method and the objective of greatest validity. These hypotheses will be comprehensive, to which new elements can be added as further data accrues. These may be data from the past as well as the present. New data should not only be capable of fitting into these hypotheses—they should be capable of being deduced from them.

Prediction

Out of the creative process just described has emerged *strategic hypotheses* that attempt to explain the individual patient. Depending on how integrated these hypotheses are, it might be said that there has now evolved a theory of the particular individual. This model views the process of prediction as vital to the process of clinical inquiry. It accepts the fact that the crucial function of the clinician in clinical inquiry is to make predictions concerning the possible courses of development in a patient. The clinician is not concerned only with prediction to the future. Statements pertaining to the present as well as the past are predictions in the sense that these facts are unavailable to the clinician and he is predicting their presence or absence in the past or in the present. But in clinical inquiry, the primary goal is to predict to the future and hypotheses are of meaning only insofar as they permit the clinician to do this. Kelly has said that clinical psychology is more than "a plot of the person's present position or a log of his past navigation" (12, p. 186). If one looks upon clinical inquiry as leading to a course of action concerning the patient, we might say that the clinical diagnostic phase is the planning stage for anticipating the direction in which the patient may move.

The model accepts the belief that predictions which are based on deductions from *strategic hypotheses* offer the greatest opportunity for validity. It is this facet of the clinical process that has been the focus of considerable discussion relative to the merits of statistical versus clinical prediction.

This model recognizes that the clinician cannot make all kinds of predictions and must be concerned with a number of issues: (1) The clinician can predict only behavior under specified conditions, i.e., contingency predictions. Such predictions are consistent with our knowledge of personality theory which indicates that consistency of human behavior cannot be considered independent of the specified conditions under which the behavior is exhibited. (2) Predictions must be expressed in the form and language of the clinician where he deals with multiple variables in describing his patient rather than in terms of rating scales which are foreign to the clinical process and destroy the configurational description which the clinician attempts to provide. (3) He can predict only propensity for behavior unless there has been an adequate study of the personality including the nature of the ego and its controls. (4) He can predict only to those situations about which there has been sufficient accumulation of knowledge and experience. Thus, in predicting to therapy

or discharge, the clinical psychologist has available a greater fund of knowledge to guide him than in predicting to success as a clinician. Actually, our knowledge is woefully lacking even in these areas. (5) Predictions should not be for events that occur at some remote time in the future. Too many factors are unknown concerning individual changes to permit predictions to events far removed from the time of the prediction. (6) All predictions should be in terms of probabilities. It is unimaginable that we can predict with certainty for there are sources of error in the clinical method as there are in all forms of inquiry. We test people in many situations but not in all. Significant variables may either not be stressed by theory and yet be very pertinent to prediction or these variables may not be revealed by the tests. In addition, theory itself is never perfect. (7) Predictions should be based upon an adequate sampling of the patient's behavior although in the practical clinical situation problems of clinical economy often do not make possible the completeness of the study demanded. (8) Predictions must be of a sort that are subject to disproof. While hypotheses may not be subject to direct disproof, predictions must be to meet the test of scientific inquiry. We must avoid those skillful predictions which are both true and false at the same time.

There has been the suggestion that the predictions that clinicians make are safe, i.e., either trivial or of general applicability. A recent study that related prediction from projective test data to clinical judgments indicate that this is *not* so (30)!

Validity of Predictions

This model requires that the psychologist follow up on his predictions, and such follow-up should be a built-in feature of clinical inquiry for if we predict wrong, no one will forget it, but if we predict right, nobody will remember it. Unless such a system of verification is built into the clinical method, it falls far short of achieving the status of a method of science. We may predict wrong, but it should be a wrong that can be self-corrective.

This model views clinical inquiry as but one part of a larger process of clinical inquiry in which are involved psychiatrists, medical practitioners, social workers, nurses and others. It is my thesis that all of these professional fields use a type of inquiry in their study of patients which may not be unlike the clinical process described herein. Each of these reaches a point in his clinical inquiry where he has prepared a communication for the other professional staff. One of the unique characters of the process of dealing with the mentally and emotionally distressed person in contemporary life is the use of the clinical team in the evaluative process. It is at a stage of inter-communication readiness that the psychological clinician communicates his findings and is able to test the validity of his predictions against other data made available by the other professions. Thus, the process of hypothesis construction and testing still continues and we may say that in true clinical inquiry it never ends until the agency's contact with the patient ceases and even here hypotheses are still re-evaluated on the basis of new data which may come to the attention of the staff.

It is important to stress the ongoing process of clinical inquiry because of recent research that has occasionally utilized the clinical report as the basis for judging the validity of predictions when this written report has been superseded by revisions in hypotheses which have not been recorded in our records. This is perhaps one of the serious limitations in our record keeping procedures which demonstrates that the scientist's recording of new developments has not been incorporated into the value system of the clinical psychologist.

The Psychological Report

Much could be written about the

failures in the clinical mode of written and verbal communication (6), but this may take us too far afield from our central theme. Suffice it to say that we have need to do much to eliminate the ambiguity, reduce the erroneous and unwarranted interpretations that others make of our communications and to maximize the usefulness of our reports. The latter is important because so much stress is being placed on accuracy of statements that we have ignored their usefulness.

Frequently, there is a semantic problem. However, more frequently, it is a failure to translate from the psychological variables emerging from the inquiry to the language that originally brought the patient for clinical investigation. Often times, there is no common language and the clinical psychologist must face the problem of sharing in the search for such a language or in recognizing that his variables, while meaningful to him, are utterly of no value to others.

NEED FOR RESEARCH

I have offered a model for clinical inquiry that I have stated achieves the status of scientific inquiry. The method of clinical inquiry herein described, like all forms of inquiry, has its errors which may result from failures anywhere in the clinical process. However, let us not fail to recognize that while this method cannot guarantee absolute certainty, that it does reduce the degree of uncertainty. How do I know? I don't! This is my conviction and the ultimate test lies in research.

There has been much talk of the evaluation of the effectiveness of various forms of predictions without sufficient consideration of how one does evaluate. The problem has arisen in the evaluation of mental health programs and Southard (31) has listed three methods of evaluation which are based on the value systems by which results are to be measured: (1) Evaluation of worth by the recipients. This would be evaluation by non-psychologists such as psychiatrists and social workers on the effectiveness of the service rendered. Here there is no specific yardstick, there is the danger of personal bias, of faulty judgment but still this is one method of evaluation that cannot be ignored. (2) Appraisal by experts who are considered to be objective. This is the method of "subjective conviction" and would be comparable to a group of psychologists subjectively evaluating a program in comparison to others. (3) The evaluation of worth by scientifically controlled methods.

We have measured the worth of the clinical method by the first two that I have listed. Our problem now is to add a more scientifically controlled type of investigation to demonstrate the efficacy of the clinical method. While Meehl (17) has collected some twenty studies demonstrating the superiority of the actuarial over the clinical method, in reality there has not yet been performed *the* crucial experiment that eliminates the serious objections to these studies (15). At a time when the challenge is being thrown to us by adherents of the actuarial method, it is not enough for us to be negatively critical about research that questions the validity of clinical inquiry. Ours must be a more constructive and responsible attitude toward designing and participating in research which will give the final answer to the raging controversy. You and I are convinced what that answer will be—the question is, are we willing to demonstrate it. My hope is that we will not be found lacking in providing the leadership for such research. After all, it is not important that we be right. What is important is that our subjective feelings about the values of our functioning *not* be accepted as scientific evidence. I share this faith with you, but we are ready now for the next stage—demonstration that this faith is valid. Statements of subjective feelings about the

worth of our clinical efforts cannot be substituted for scientific evidence. History is too replete with examples of the final rejection of mistaken convictions.

References

1. Baughman, E. E. Rorschach scores as a function of examiner differences. *J. proj. tech.*, 1951, *15*, 243-249.
2. Cofer, C. N. An analysis of the concept of "clinical intuition." In G. A. Kelly (Ed.), *New methods in applied psychology*. College Park, Maryland: Univ. of Maryland, 1947.
3. Cohen, M. R. and Nagel, E. *An introduction to logic and scientific method*. New York: Harcourt, Brace, 1934.
4. Deutsch, M. Field theory and projective techniques. *J. proj. tech.*, 1954, *18*, 427-434.
5. Holt, R. R. The thematic apperception test. In H. H. and G. L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951.
6. Holzberg, J. D., Alessi, S. L., and Wexler, M. Psychological case reporting at psychiatric staff conferences. *J. consult. Psychol.*, 1951, *15*, 425-429.
7. Horst, P. et al. *The prediction of personal adjustment*. New York: Social Science Research Council, 1941.
8. Hutt, M. L. Assessment of individual personality by projective tests. *J. proj. tech.*, 1951, *15*, 388-393.
9. Hutt, M. L. et al. The effect of varied experiment "sets" upon Rorschach test performance. *J. proj. tech.*, 1950, *14*, 181-187.
10. Jahoda, M., Deutsch, M., and Cook, S. W. *Research methods in social relations*. New York: Dryden, 1951.
11. Joel, W. The interpersonal equation in projective methods. *Rorsch. Res. Exch.*, 1949, *13*, 479-482.
12. Kelly, G. A. *The psychology of personal constructs*. New York: Norton, 1955.
13. Lord, E. Experimentally induced variation in Rorschach performance. *Psychol. Monogr.*, 1950, *64*, No. 10.
14. Maslow, A. H. Problem-centering vs. means-centering in science. *Phil. Sci.*, 1946, *13*, 326-331.
15. McArthur, C. C. Clinical vs. actuarial prediction. In *Invitational conference on testing problems*. Princeton, N. J.: Educational Testing Service, 1955.
16. McClelland, D. C. Toward a science of personality psychology (In press).
17. Meehl, P. E. *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press, 1954.
18. Meehl, P. E. Wanted—A good cookbook. *Amer. Psychol.*, 1956, *11*, 263-272.
19. Miller, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.*, 1953, *48*, 367-375.
20. Murray, H. A. Research planning: A few proposals. In S. S. Sargent and M. W. Smith (Eds.), *Culture and personality*. New York: Viking Fund, 1949.
21. Rapaport, D., Gill, M. and Schafer, R. *Diagnostic psychological testing*. Chicago: Yearbook Publishers, 1945.
22. Rapaport, D. The theoretical implications of diagnostic testing procedures. In *Congres International de Psychiatrie*. Paris: Hermann and Cie, 1950.
23. Ross, S. Sucking behavior in neonate dogs. *J. abnorm. soc. Psychol.*, 1951, *46*, 142-149.
24. Sanders, R. and Cleveland, S. E. The relationship between certain examiner personality variables and subjects' Rorschach scores. *J. proj. tech.*, 1953, *17*, 34-50.
25. Sarason, S. B. *The clinical interaction*. New York: Harper, 1954.
26. Schachtel, E. G. Subjective definitions of the Rorschach test situation and their effect on test performance. *Psychiat.*, 1945, *8*, 419-448.
27. Schafer, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune and Stratton, 1954.
28. Shneidman, E. S. *Thematic test analysis*. New York: Grune and Stratton, 1951.
29. Sipola, E. and Taylor, V. Reactions to ink blots under free and pressure conditions. *J. Pers.*, 1952, *21*, 22-47.
30. Sohler, D. T., Holzberg, J. D. et al. The prediction of family interaction from a battery of projective techniques. *J. proj. tech.* (in press).
31. Southard, C. G. Symposium on the evaluation of community mental health programs. Reported in *Evaluation in Mental Health*. Washington, D. C.: U. S. Dept. of Health, Education and Welfare, 1955.
32. Stein, M. I. Personality factors involved in the temporal development of Rorschach responses. *Rorsch. Res. Exch.*, 1949, *13*, 355-414.
33. Tomkins, S. S. *The thematic apperception test*. New York: Grune and Stratton, 1947.
34. White, R. W. What is tested by psychological tests? In P. H. Hoch and J. Zubin (Eds.), *Relation of psychological tests to psychiatry*. New York: Grune and Stratton, 1951.

Received July 2, 1957.

Objectifying the Subjective—A Methodological Approach to the TAT [1]

IRA FRIEDMAN
Cleveland Receiving Hospital and State Institute of Psychiatry

Experimental work with projective techniques presents many problems to the research worker. One of the more difficult problems is related to the discrepancy between the objective, atomistic approach which usually characterizes this type of research, and the subjective, holistic approach which is representative of the manner in which projective instruments are used in clinical practice. On the one hand validity seems to be sacrificed for the sake of reliability while, on the other hand, reliability is sacrificed for the sake of validity. Many attempts are being made to harness subjectivity and holism in the service of reliable, objective research. The present study represents an attempt in this direction. More specifically, it presents a methodological approach utilized in the analysis of the Thematic Apperception Test (TAT) for research purposes and the reliability of judges employing this method.

In reference to research with the TAT, there has been an emphasis upon frequency of theme tabulations (4, 10) and objective scoring approaches (1, 3, 6, 8). These methods seriously limit the subjective, holistic appraisal of protocols and may well overlook important qualitative features. Perhaps a step toward the solution of this difficulty is touched upon by Eron (5) who states, "The use of rating scales, however, affords an opportunity to combine qualitative, clinical judgments and a more or less rigorous quantification. It is possible for the ratings to be as subjective and holistic as necessary, but it is essential that the criteria be verbalized so that the method can be communicated and subsequent raters can utilize the scales with comparable results." In the present study, Eron's recommendations have been modified in one important respect. Separate rating scales for isolated variables have been avoided. Instead, one rating scale with a number of variables is employed with the aim of approaching more closely the description of personality, in this case characteristics of the TAT hero, as a Gestalt. In addition, the criteria utilized for rating are broad and generalized without making the rating procedures so specific that it leaves the individual rater little freedom for subjective appraisal. The criteria for rating protocols in the present study is a compromise between the objective and subjective approaches and is effected by utilizing, as rating procedures, many of the clues and methods utilized in the clinical application of the TAT.

METHOD

Eighty statements were prepared, representing the TAT hero characteristics that appeared in a random sample of TAT stories and in the common themes to TAT cards as given by Stein (12). The statements were derived from stories to Cards 1, 3BM, 6BM, 7BM, and 14, and reflect the traits and attributes, feelings and emotions, experiences, tendencies toward behavior, goals and orientations that people generally attribute to their TAT heroes. These statements are presented in order to give the reader an orientation to the variables which were considered.

[1] The basic result of this research is presented elsewhere (7). The present article represents an elaboration of the method as a demonstration of one approach to the TAT for research purposes.

List of Statements

1. ambitious
2. works hard to achieve a desired goal
3. discouraged over lack of success
4. capable
5. believes in business before pleasure
6. an independent person
7. gets very interested in things
8. dislikes inactivity
9. envies others
10. has a good imagination
11. has difficulty sleeping
12. a serious person
13. gets drunk
14. gambles
15. stubborn
16. sometimes felt like leaving home
17. tries to comfort people in their sorrow
19. thrifty
19. concerned about appearance
20. understanding of peoples' problems
21. tells people what to do
22. enjoys the beauty of nature
23. has been a disappointment to father in some respects
24. thinks about travelling in far off places
25. takes good care of possessions
26. learns from experience
27. sensitive
28. has had unhappy love experiences
29. becomes discouraged easily
30. dislikes telling sad news
31. feels blue and downhearted
32. worries a great deal
33. sometimes thinks about suicide
34. displeased with home life in some respects
35. often feels tired
36. gets over disappointments quickly
37. sometimes feels like crying
38. upset by things which later seem unimportant
39. enjoys being alone
40. gets depressed when alone
41. irritable
42. has good control over emotions
43. feels bitter toward people
44. tries to get even with people
45. rarely expresses anger
46. finds it difficult to control temper
47. thinks about enjoyable things
48. thinks how nice it would be to be a success
49. gets into trouble
50. ashamed after doing something wrong
51. looks at the bright side of things
52. fears doing the wrong thing
53. finds it difficult to make decisions
54. daydreams a good deal
55. concerned about the future
56. would never steal under any circumstances
57. forced to do unenjoyable things
58. resents being forced by parents to do things
59. obeys parents
60. does what other people advise
61. often wronged by other people
62. was punished by father
63. was punished by mother
64. has felt angry at father
65. has felt angry at mother
66. spoiled by parents
67. sometimes disagrees with mother
68. has been a disappointment to mother in some respects
69. argues with mother when there is a difference of opinion
70. feels close to mother
71. liked to talk over troubles with mother
72. has confidence in mother's advice
73. as efficient as father
74. rarely asked mother's permission to do things
75. follows father's advice
76. feels close to father
77. dislikes being told what to do
78. respects father
79. helped by father's guidance
80. treated like a child by father

The Q-sort rating technique (13) was adopted to give a single composite description of any individual's TAT heroes. It was the task of the judges to sort these 80 statements into nine piles approximating a normal distribution ranging from those statements which were most characteristic of the TAT hero to those which were least characteristic.

A set of instructions, more properly called suggestions, were drawn up to provide some framework whereby the raters could handle problems of sorting in much the same manner. These instructions represent an attempt to achieve some uniformity without sacrificing the more global, clinical approach for the sake of reliability. It will be noted that depth interpretation was reduced to a minimum by placing primary emphasis upon manifest content. The emphasis upon manifest content does not mean that subjectivity was eliminated. In fact the decision to avoid a complicated scoring manual was motivated by the desire to allow clinical judgment and impressionism to

operate more freely than would be possible if the rater were restricted by rigid scoring rules. While this method sacrifices some degree of reliability, it maintains much more of the unique features of the individual protocol than scoring schemes or frequency of theme tabulations.

INSTRUCTIONS FOR Q-SORT RATINGS FROM TAT PROTOCOLS

A. In the five stories there will usually be several outstanding qualities of the hero or of the hero's behavior. These belong in the extremes even when it is necessary to employ a slight degree of inference. For example, we can score the statement "feels close to mother" as high even though it is not expressed, if the mother is pictured as understanding, if the subject likes to take troubles to her, if she helps him, if he is concerned about her, and if there are no negative feelings expressed. These outstanding characteristics can frequently be chosen by the length of the story, the intensity of affect, the embodiment of the essence of the particular story, the frequency of appearance, and numerous qualitative indications.

B. Directly given statements such as tiredness, daydreaming and the like should not be weighted too highly when they are not particularly stressed. Although items which are directly given receive higher weighting than items of the same intensity but not directly given, the rater should always keep in mind the necessity of estimating the strength or intensity as expressed in the TAT.

C. The more *frequent* and the more *unusual* a quality, characteristic, or behavior of the hero, the higher it should be weighted.

D. Omissions should be considered. For example, if the stories reflect particularly stressful scenes where we would normally expect the expression of anger, then we would weight the item "rarely expresses anger" highly (if anger is omitted). Contrariwise, reactions which are out of keeping with the expressed situation are given additional weight.

E. Where a certain behavior in the stories implies, rather directly, a certain feeling or attitude, this may be weighted in accordance with the intensity of the expressed behavior (see example in A, also, crying may imply depression or shame or lack of emotional control, etc., depending upon the rest of the story).

F. Because of inconsistencies in personality, differences in the stimulus properties of the card, etc., contradictory themes frequently appear in the TAT. It is similarly possible to have contradictions in any of the rather broad Q-sort categories. The contradictory elements should be considered separately and weighted appropriately, e.g., a person may be both elated and depressed, energetic and inactive. Sometimes the slightly different wording on the same general variable may prove useful, e.g., a person may be ambitious but does not work hard to achieve a desired goal. However, when there is only one card pertaining to a variable where contradictory conceptions have been expressed, determine which of the polarities seem most important but do not weight it as highly as an item where no contradictory conceptions have been expressed.

G. Where several themes are expressed on one card, the separate stories should be considered and weighted appropriately.

H. When it is necessary to ask yourself the question "would a person like this be such and such" there is evidence that the variable under consideration is not really expressed in the TAT and that an increasing degree of inference is being employed. It is true that it will be necessary to do this to fulfill the requirements of the Q-sort but these items should not be weighted as highly as those which are more directly obtainable from the TAT protocol.

I. The middle categories of the Q-

sort should consist, mainly, of those items which are indeterminate from the TAT stories. However, if the majority of the items *fit* the TAT hero, it may be necessary to place the indeterminate items on the least characteristic side.

J. One should be careful to remember that we are sorting for the qualities of the TAT hero and not for qualities of the subject. This is sometimes a very difficult distinction to make. For example, if the subject gives several alternative stories or cannot decide upon an outcome, we often infer that he finds it difficult to make decisions. However, this is not the same as saying that the TAT *hero* finds it difficult to make decisions. To come to that type of conclusion for the purposes of this research we would expect the story hero to be confronted by a situation where he (the hero) cannot make a decision.

Five judges were selected, all of whom had had experience with the TAT. Each judge was given a sheet of instructions for rating, the rating procedures were discussed, and the judges sorted one sample protocol with the author to get a general impression of the task. Ten records, both normal and deviant, were selected at random. The author rated each of the 10 records and secured two other judges' ratings for each record so that there were three ratings or sorts for each record.

Results

Since three judges rated each record, judge A's sort could be correlated with judge B's and C's, and judge B's and C's with each other. The three reliability coefficients (AB, AC, BC), each reflecting the degree of agreement between two judges in characterizing the TAT hero, were computed by means of the product-moment correlation formula suggested by Cronbach (2) for use when the scores of all persons have the same mean and variance as in the case of Q-sort data. The 30 reliability coefficients obtained from the three correlations for each of the 10 records, range from .37 to .88 with a mean correlation of .72 and a median of .75. When computing the average correlation by the method of z-transformation of a correlation coefficient, the reliability figure was .74. The frequency distribution of inter-rater reliability coefficients is presented in Table I.

Table I—Frequency Distribution of Inter-Rater Reliability Coefficients

Reliability Coefficient	Frequency
.80 – .89	4
.70 – .79	19
.60 – .69	4
.50 – .59	1
.40 – .49	1
.30 – .39	1

Discussion

The obtained correlations compare favorably with other studies reported in the literature (1, 9, 11) dealing with similar types of ratings, and indicate that judgments on this type of material can be reliably made. While studies with more objective types of data, or more objective scoring manuals are often capable of achieving a higher degree of reliability, the present method is offered as an attempt to deal with the test in a clinical, holistic way and still maintain an acceptable degree of reliability.

In the present study, the judge's task was to characterize the story hero in accordance with the major purposes of the larger study in which this method was utilized (7). This is a limiting factor of the reported investigation as a characterization of the story *teller* rather than the story *hero* is more closely allied to the application of the TAT in the clinical setting. The methodological approach appears equally applicable to the story teller as well as the story hero, and, as such, offers many possibilities for further research which can im-

prove our understanding of what can and cannot be done with the TAT, as well as to improve methods for doing it. In particular, the statements can be changed so that many different variables, or trait universes may be investigated—psychoanalytic character traits, methods of interpersonal relatedness, Murray's needs, etc. By adapting rating instructions in a variety of ways, it is possible to learn something about the reliability or unreliability of judgments from the TAT, concerning such things as estimating self concept, predicting behavior, characterizing core conflicts, assessing attitudes toward psychotherapy, judging feelings surrounding peer, sibling, and authority relationships, etc. In addition, by comparison with outside criteria, validity can be readily assessed or, at least, reliability of judgments from two different sources of data.

SUMMARY

The present research is a reliability study of a method of TAT analysis for research purposes. This method represents an attempt to employ subjectivity and holism by utilizing a) a single multivariable rating scale (Q-sort) rather than separate univariable rating scales, and b) a broad descriptive set of rating instructions making explicit many of the subjective clues used in the interpretation of the TAT in clinical practice rather than a detailed rating or scoring manual which leaves little freedom for subjectivity and impressionism. The average correlation of inter-rater reliability coefficients was .74. The results tend to suggest that this approach is capable of yielding acceptably reliable results. The suggestion that different rating instructions and different statements may be employed in the investigation of a variety of problems is made.

REFERENCES

1. Aron, B. *A manual for analysis of the Thematic Apperception Test.* Berkeley, California: Willis E. Berg, 1949.

2. Cronbach, L. J. Correlations between persons as a research tool. In O. H. Mowrer (Ed.), *Psychotherapy theory and research.* New York: Ronald, 1953, Pp. 376-388.

3. Dana, R. H. An application of objective TAT scoring. *J. proj. Tech.*, 1956, *20*, 159-163.

4. Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol., Monogr.*, 1950, *64* (Whole No. 315).

5. Eron, L. D. Some problems in the research application of the Thematic Apperception Test. *J. proj. Tech.*, 1955, *2*, 125-129.

6. Fine, R. Manual for a scoring scheme for verbal projective techniques. *J. proj. Tech.*, 1955, *19*, 310-316.

7. Friedman, I. Phenomenal, ideal, and projected conceptions of self. *J. abn. soc. Psychol.*, 1955, *51*, 611-615.

8. Goodman, M. An indirect validation of a TAT scoring manual. *J. clin. Psychol.*, 1952, *8*, 149-154.

9. Harrison, R., and Rotter, J. B. A note on the reliability of the TAT. *J. abn. soc. Psychol.*, 1945, *40*, 97-99.

10. Rosenzweig, S., and Fleming, E. E. Apperception norms for the TAT. *J. Pers.*, 1949, *17*, 483-503.

11. Sanford, R. N., et al. Physique, personality and scholarship. *Monographs of the Society for Research in Child Development*, 1943, *8*, p. 705.

12. Stein, M. I. *The Thematic Apperception Test: An introductory manual for its clinical use with adult males.* Cambridge, Mass.: Addison-Wesley, 1948.

13. Stephenson, W. *The study of behavior.* Chicago: University of Chicago Press, 1953.

Received May 6, 1957

Responses of School-Children to Human and Animal Pictures[1]

KENJI FURUYA
Department of Psychology, Tokyo University of Education, Japan

PROBLEM

In his C.A.T. (Children's Apperception Test), Bellak (1, 2) used animal pictures because he believed that children identify themselves much more readily with animal than with human pictures. This assumption by Bellak was tested by Bills (4), who conducted an experiment in which forty-eight children told stories about both T.A.T. (human) cards and animal cards, and it was found that almost all the children told longer stories to the animal cards than to the human cards. On the other hand, Light (5) compared five C.A.T. cards with five T.A.T. cards. In this experiment, using 75 fourth and fifth graders with a mean age of nine years and eight months, five criteria out of six yielded significant differences which suggested the superior productivity of T.A.T. (human) cards to C.A.T. (animal) cards. Biersdorf and Marcuse (3) pointed out that in comparing the productivity of animal and human pictures it would seem necessary to consider the nature and the comparability of the situations portrayed in the two sets of pictures. When two sets of dissimilar pictures were employed, it would be difficult to determine which difference was responsible for the result obtained: situations or the use of animal and human figures. Biersdorf and Marcuse chose six cards, Nos. 1, 2, 4, 5, 8, and 10 from the ten cards in Bellak's C.A.T., and six corresponding pictures utilizing humans. The scenes, situations, sizes and the shadings of each sets (human and animal) were made as nearly equivalent as possible. Thirty first-graders with a mean age of seven, were asked to tell stories in response to these two series of animal and human pictures. There was no significant difference in productivity between six animal pictures and six human pictures. Using the same method as Biersdorf and Marcuse (3), Mainord and Marcuse (6) experimented with emotionally disturbed children and found that one criterion out of six yielded a significant difference which suggested the superior productivity of human pictures.

Since Bellak believes that the C.A.T. is suited for children from three to ten and the T.A.T. for adolescents and adults, Bellak's assumption seems to suggest that the superior productivity of animal pictures to human pictures decreases relatively as the children grow older. The present study was thus designed to answer the following questions:

1. With different criteria and age-groups (from those of Biersdorf, Mainord and Marcuse), what kind of difference will be found between the productivity to animal pictures to human pictures, equivalent in scenes and situations?

2. Is there any tendency toward relative decrease of productivity to animal pictures compared with human pictures as the children grow older?

METHOD

(a) *Test Materials*

Test materials consisted of the two sets of six animal cards and six human cards which were utilized by Biersdorf and Marcuse (3) and Mainord and Marcuse (6). Only the pictures showing the scene of toileting, similar to No. 10 card of Bellak's C.A.T., were partially modified for

[1] The author of this study is indebted to Dr. F. L. Marcuse of Washington State College for the use of the negative photo-plates of his H.A.T.

Japanese children.[2]

(b) *Subjects*

Seventy-two children served as Ss. They were divided into the following three groups. Group 1 consisted of 24 first-graders; their ages ranged from 6 years 4 months to 7 years 3 months, with a mean age of 6 years 9 months. Group 2 consisted of 24 fourth-graders; their ages ranged from 9 years 7 months to 10 years 4 months, with a mean age of 9 years 11 months. Group 3 consisted of 24 sixth-graders; their ages ranged from 11 years 6 months to 12 years 5 months, with a mean age of 11 years 10 months. Each of the three groups consisted of 12 boys and 12 girls. All children were selected randomly.

(c) *Procedure*

Each of the three groups was divided randomly into two smaller groups, A-group and B-group. Each of the three A-groups and the three B-groups consisted of 12 Ss, 6 boys and 6 girls. There were two sessions two weeks apart. The members of A-groups were presented with the six animal cards in the first session and with the human cards in the second session; the members of B-groups were presented with the human cards in the first session and the animal cards in the second. The order of presentation of the cards varied from S to S, but was constant for each S for the two sessions.

Group 1 was administered the test individually. An attempt was made to establish rapport by means of an individual standardized interview. The stories which the child told in response to cards were written down by the experimenter.

Groups 2 and 3 were administered the test as a group. Six children from each of groups 2 and 3, consisting of three A-group members and three B-group members selected randomly, were tested at a time. Each child was seated in a child's chair at a child's desk, at a distance of about two meters from each other. After distribution of pencils and paper, the children were asked to write stories in response to pictures which the experimenter was going to show them. Then each child was given his card in the previously determined order of the presentation. The experimenter encouraged Ss to make stories as long as possible. Eight minutes after the presentation of the cards, the experimenter said, "You will finish your story after two more minutes, and now you will write the outcome of your story." After ten minutes the experimenter collected the cards. A minute later a new card was distributed to each S upside down. Most of the children finished their stories in the allotted time. Subsequent cards were treated in the same way as the first one. A five minutes recess was given between the third card and the fourth.

RESULTS

When the test was over, most children of all groups said that the test was very interesting. No card was rejected by any child. One hundred and forty-four animal stories and 144 human stories were told by each of groups 1, 2 and 3. The total number of the stories was 864. These stories were analyzed on the following eight criteria:

1. Response time in seconds, the length of time between the presentation of the cards and the S's first word in response to the card.
2. Time of total response in seconds, the length of time between the S's first word and his last word in response to the card.
3. Number of "Bunsetsu" (similar to words).
4. Number of characters mentioned.

[2] Since most Japanese children do not know the American or European style toilet and mistake it for something else, there occurs considerable discrepancy between the responses of children who know the American style toilet and of those who do not know it. Therefore, the Japanese style toilet was substituted for the American style.

5. Introduced figures.
6. Expressions of feeling.
7. Expressions of significant conflict.
8. Outcome, if present.

Two criteria, 1 and 2, were not used in the case of groups 2 and 3, because they were examined in groups. To test the reliability of the analysis, the test-retest method was utilized. The author analyzed the same stories twice with one month interval. An agreement of 94% was obtained.

Table I shows the comparison between the responses to the six animal pictures and the six human pictures. In Group 1, only one criterion (outcome) yielded a significant difference. In Group 2, two criteria (feeling and conflict) yielded significant differences. In Group 3, one criterion (feeling) yielded a significant difference. All of these four statistically significant differences are of the kind which suggests the superior productivity of human pictures to animal pictures.

Fig. 1, 2, 3, 4, 5 and 6 show the developmental aspects of the productivity of animal and human pictures. These figures would suggest that the relative productivity of animal and human pictures does not decrease as the children grow older from six years to eleven.

In Fig. 3, the percentage of animal stories containing introduced figures are apparently always more than those of human stories containing introduced figures. When the two average percentages of the three groups were calculated, the percentage of the stories containing introduced figures was 31.5% for animal stories, and 27.5% for human stories. But this difference was not statistically significant.

Discussion

In this experiment, one criterion

Table I—Comparison Between the Mean Responses of Children (N-72) to Six Animal Pictures and to Six Human Pictures.

Criterion	Group	Animal Cards Mean	Animal Cards S.D.	Human Cards Mean	Human Cards S.D.	t	p
1. Response time per card in sec.	1	6.06	5.52	6.56	6.50	.69	....
2. Time of total response per card in sec.	1	273.8	160.0	271.8	134.8	.11	....
3. Number of "Bunsetsu" per card	1	61.74	50.2	58.0	45.6	.69	....
	2	67.65	25.4	68.14	27.3	.16	....
	3	80.64	28.6	79.94	29.8	.15	....
4. Number of characters mentioned (in the picture) per card	1	2.79	0.91	2.80	.077	.10	....
	2	2.66	0.94	2.79	0.89	.10	....
	3	2.55	0.86	2.64	0.93	.09	....
						X^2	P
5. Percentage of stories containing introduced figures	1	36.8		30.5		2.07	....
	2	23.0		20.8		1.14	....
	3	34.7		31.2		.64	....
6. Percentage of stories containing expressions of feeling	1	42.3		49.9		2.34	....
	2	68.0		75.0		3.85	.05
	3	72.2		81.2		5.83	.02
7. Percentage of stories containing expressions of significant conflict	1	28.4		30.5		0.31	....
	2	6.5		13.9		9.30	.01
	3	18.8		19.3		.05	....
8. Percentage of stories having definite outcome	1	13.2		19.4		6.23	.02
	2	27.7		25.7		.44	....
	3	50.1		50.0		.05	....

FIGURES SHOWING THE DEVELOPMENTAL ASPECTS OF THE COMPARISON BETWEEN THE PRODUCTIVITY OF ANIMAL CARDS AND HUMAN CARDS

Animal Cards — black bar

Human Cards — white bar

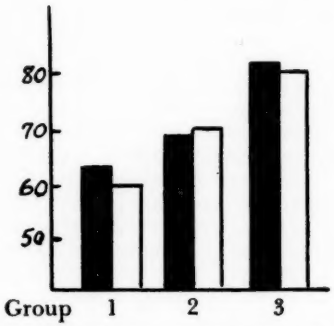

FIG. 1 The average number of "Bunsetsu" (words) per cards

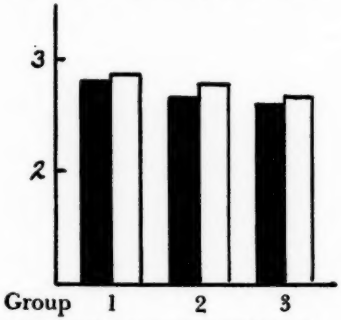

FIG. 2 The average number of characters mentioned (in the picture) per cards

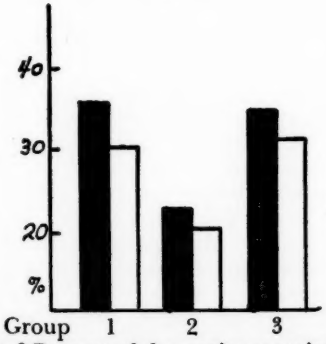

FIG. 3 Percent of the stories containing introduced figures

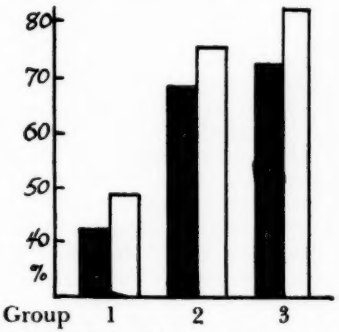

FIG. 4 Percent of the stories containing the expressions of feeling

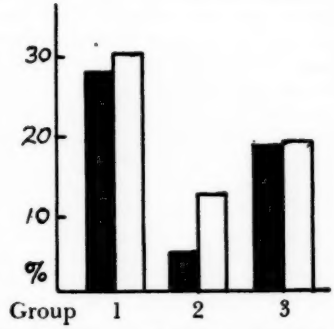

FIG. 5 Percent of stories containing expressions of significant conflict

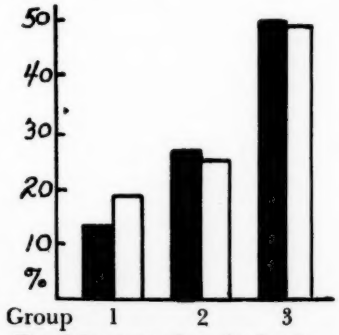

FIG. 6 Percent of stories having definite outcome

out of eight in Group 1 (6-year-olds) two criteria out of six in Group 2 (9-year-olds), and one criterion out of six in Group 3 (11-year-olds) yielded significant differences (at the 5% level or better), suggesting the superior productivity of human pictures. All of these criteria which yielded significant differences are different from those of Biersdorf and Marcuse (3) and Mainord and Marcuse (6). Developmentally, the results obtained did not show any tendency to indicate that the relative productivity of animal and human pictures, decreases or increases as the children grow older in the year range of the Ss in this experiment, and the criteria utilized.

These results contradict Bellak's assumption, and seem to suggest that children identify themselves a little more readily with human pictures than with animal pictures.

To answer fully the question raised in this study, further experimentation with pictures utilizing other situations and scenes, other age-groups, other criteria, and with normal and disturbed children are required. Furthermore, the author of this study wishes to point out that most of the animal pictures, that have been used, were "personified or human-like" animal pictures. Therefore, in considering whether or not children identify themselves more readily with animal pictures or human pictures, it would seem necessary to compare "real or natural" animal pictures with human pictures. There might be some differences between "personified" animal pictures and "real" animal pictures.

Summary

Twenty-four first-graders (Group 1), 24 fourth-graders (Group 2) and 24 sixth-graders (Group 3) were asked to tell stories in response to six animal pictures and six human pictures, equivalent in scene and situations. This was done in order to determine whether there was a significant difference between the productivity of stories told to animal pictures and human pictures, and whether there was any tendency to indicate that the relative productivity of animal pictures compared with that of human pictures decreases as the children grow older. The results obtained were as follows:

One criterion in Group 1, two criteria in Group 2 and one criterion in Group 3 yielded significant differences which suggest the superior productivity of human pictures. As to the developmental aspects of the problem, the data obtained did not show any tendency for the relative productivity of animal and human pictures to change with age.

References

1. Bellak, L. *The T.A.T. and C.A.T. in clinical use.* New York: Grune & Stratton, 1954.

2. Bellak, L., and Bellak, S. *Children's Apperception Test.* New York: C.P.S. Company, 1949.

3. Biersdorf, K. R., and Marcuse, F. L. Responses of children to human and animal pictures. *J. proj. Tech.*, 1953, *17*, 455-459.

4. Bills, R. E. Animal pictures for obtaining children's projection. *J. clin. Psychol.*, 1950, *6*, 291-293.

5. Light, B. H. Comparative study of a series of T.A.T. and C.A.T. cards. *J. clin. Psychol.*, 1954, *10*, 179-181.

6. Mainord, F. R., and Marcuse, F. L. Responses of disturbed children to human and animal pictures. *J. proj. Tech.*, 1954, *18*, 475-477.

Received April 16, 1957

Rorschach: Some Comments on Predicting Structured Behavior from Reactions to Unstructured Stimuli[1]

ARNO GRUEN
Northside Center for Child Development
and
Cornell University Medical College

Much work has been directed toward an understanding of the stimulus and task situation presented by the Rorschach test. Fortier (1), Lazarus (9), and Schachtel (16) have been concerned with isolating the function of color and understanding the nature of the color response, Lord (11) and Gibby (2) have concerned themselves with the effect of the examiner on Rorschach performance. Hutt, et al. (5) have considered the effect of "set" on Rorschach performance. Schachtel (17) has written on the nature and quality of the human movement response.

In all this work the analysis is in terms of factors presumed to operate in the test or in the test situation. Nowhere is the point of departure an analysis of the ambiguous *nature* of the stimulus situation of the test itself and its role in the attempt to predict life behavior from reactions to the Rorschach cards.

Though the nature of the test situation is verbalized frequently enough, it has perhaps become so implicitly accepted that the question of its meaning has never been posed explicitly. Clinicians speak of it as an ambiguous or unstructured stimulus situation (6, 12). As such it is the way in which the examinee handles the unstructured stimuli provided by the cards that is taken as the basis for prediction of the individual's behavior.

The implication is that we can infer behavioral structure from behavior elicited in a situation whose dimensions are purposely vague and unstructured. In other words, the assumption underlying projective theory here is that behavior in an unstructured situation tells something of behavior in real life situations.

The rationale for the necessity of utilizing unstructured material is relatively simple. Since individuals will behave in ways which they feel are socially acceptable it becomes necessary to frustrate this tendency by providing a novel stimulus situation (Rorschach cards) to elicit totally new behavior. This "new" behavior (the Rorschach responses) is then used as a basis for predicting the individual's real life behavior and motivation.

The question that presents itself is: If the test elicited behavior is a function of a totally new and unstructured situation, introduced to prevent the person from behaving in what he thinks are socially acceptable ways, what kind of real life behavior can such behavior predict? Does it predict how the individual really behaves to others and with himself? But the test is given in the first place to overcome his acceptable picture!

Is it that from the unstructured situation and the individual's response to it that we can predict how he would behave if he were to find himself in an analogous real life unstructured situation?

Several questions which we cannot, however, take up at this time, present themselves at this point. To what extent is the Rorschach really a completely unstructured stimulus situation? Ranzoni, Grant & Ives' (13) work illustrates that some of the stim-

[1] The writer is very indebted for critical comment and discussion to the following persons: Kenneth B. Clark, Mamie P. Clark, Pauline Fernandez, and Daniel S. Lehrman.

uli constellations within the cards have more organization than others. Thus, Rorschach responses must be qualitatively different in terms of whether or not they are responses to more or less structured stimuli situations. That is to say that the extent to which they are or are not imposed on the perceiver through the organization inherent in the stimulus field would bring very different aspects of the perceiver's contribution to the act of "perception" into the situation.

Thus the Rorschach could also be thought of as providing the subjects with the *choice* of responding at different levels of perception. This, of course, raises the problem of *interpreting the choice.* That is, is it a conscious choice, is it imposed upon the person by the stimulus situation in terms of the prior experiences the individual has had with various sensory aspects of the stimulus field (see for example Schachtel (16)), or is it a question of repression, intelligence, negativism or combinations of these?

However, for the purposes of the present paper we are confining ourselves to the examination of the Rorschach as an unstructured stimulus situation, since some of the most important cornerstones of Rorschach analysis, such as movement and color reactions, reflect the process of imposing structure by the perceiver on the ambiguous aspects of the stimulus material.

In 1949[2] the author experimented with predictions from Rorschach protocols of perceptual judgments of a falling object both within a frame of reference and with a much reduced frame of reference.

A series of 36 cards were shown to each of 24 undergraduate students under two conditions. On each card the picture (in silhouette) of a man tied on a chair had been printed. Each succeeding picture showed the chair tilted backward an additional half degree. Under the first condition the pictures, each printed on a white 8½ x11 cardboard, were shown 15 inches from the subject's face, exposed one at a time, with the subject's head fixed in a straight clamp. The subject's task was to say "stop," when he felt that that picture was exposed in which the man was seen as just starting to fall over backward.

The same judgment had to be made again with the subject viewing the whole sequence of 36 positions at one time, arranged consecutively on a large cardboard. In this way it was felt that a framework for judging was introduced.

Two judgments were taken under condition 1, one hour apart. The product moment correlation was + .82. In a repetition of the perceptual part of this experiment at the College of the City of New York with 40 subjects, the correlation for two judgments taken four weeks apart was +.94. The correlation for the second condition, (administered twice in the second experiment) when all positions of the figure were simultaneously exposed, was +.83. All correlations were significant.

In terms of Rorschach's theory (14) it was felt that Rorschach signs concerning empathic involvement would correlate with degree of movement permitted the silhouetted figure. Accordingly a chi-square 2x2 contingency table showed that the probability of the relationship of more tilt[3] to intra-tensiveness (M larger than C), and of less tilt to extra-tensiveness (C larger than M) was better than chance at the 1% level when judgment of the tilting figure under the first condition was considered.

The contingency coefficient here is +.59 with a P.E. of .109. On the other hand these Rorschach categories bore no relationship to judgment of falling under the second condition of the ex-

[2] This work was carried out in the Graduate Psychology Department of New York University. The Department generously supplied the funds for building the apparatus.

[3] Using the Mean of judgments as the cut-off point.

periment, when a full framework of all tilting pictures was simultaneously available.

The mean in degree of judgment under the first condition for the New York University students was 11 seconds, with a S.D. of 4.47. For the second condition the mean was 18.66 seconds with a S.D. of 6.18. The difference between the means is significant at better than the 1% level.

In the repetition of the experiment at C.C.N.Y. the mean for the first condition was 11.60, S.D. 4.39. This mean did not differ significantly from the N.Y.U. group.

The mean for the second condition here was 15.42, the S.D. 4.30. The probability of this mean being significantly different from that of the N.Y.U. group is rated at the 20%-30% level. The two C.C.N.Y. means are significantly different from each other beyond the 1% level.

Thus the two conditions of judgment of tilt apparently reflect consistent differences in the nature of the judgment situation. In one, judgment must be made on the basis of reduced cues, in the other a whole array of other tilting positions is visible and thus the observer has a consistent frame of reference before him. Under the first condition the "Anhaltungspunkte" would appear to be primarily subjective (13) and thus much more akin to a situation without objective structure. In the second we have a much more narrowly defined situation, with a great array of cues for judging relative position.

What is of interest here is that the relation to Rorschach variables of the same perceptual judgment is significant if the judgment occurs under less structured conditions, but is without significance when the judgment is made under relatively more structured conditions.

Thus we have a situation where the Rorschach predicts behavior under unstructured conditions but not when the situation becomes more organized.

This in some measure would seem to approximate the dilemma that clinicians so often come across (5, 7) where a "psychotic" record obtained within a mental institution reflects the psychotic behavior of the patient, while a similarly "psychotic" record taken from a non-hospitalized individual does not. The "validity" of the Rorschach interpretation in the first instance might possibly arise out of the equivalence that the Rorschach stimulus situation and the environment have for the patient under consideration. To the psychotic patient the social orderedness of things is not available. Each real life stimulus situation requires a new integration and new appraisal because of his difficulty in perceiving it in a well ordered socially standardized manner. Thus his manner of approach to integration on the Rorschach may very likely reflect in some measure what he also does in real life. However, where the individual perceives his environment in terms of a socially determined structure, this equivalence would diminish, depending on the degree to which his perception of the everyday world is socially integrated. These considerations suggest that the easy transition from behavior in one area to predicting behavior in another be checked by a thorough evaluation of how structured the world is for the person under examination.

Hertzman and Pearce in their highly penetrating paper on the personal meaning of the human movement response (4) and also Kornreich (8) and to some extent Wittenborn (19) have pointed out that prediction from Rorschach data to behavior cannot be univariant and must take cognizance of the structure of the individual's life surroundings and his reactions to it in order to be most meaningful. We must know something about an individual's way of structuring reality before we can readily assume a parallelism that would enable us to jump from behavior on an unstructured task to behavior that occurs under conditions that have structure and

also impose a greater degree of reality consideration.

The problem entailed in making such inferences has been passed over for too long by simply speaking of "potentials" (7) when face validity was evidently non-existent. Thus it may be that Rorschach performance rather than being directly related to social behavior reflects behavior at the levels of phantasy and feeling far removed from "reality" considerations, comparable to the primary process of Freud (10). As such, they may have importance for understanding the primitive and magical nature of feelings behind our logical and socialized exterior. However, unbridled illogical construction on the Rorschach may not permit inference about psychosis simply because psychotics show such construction. It might denote that for others, besides psychotics, the Rorschach permits a free flow of subjective "illogical structuring"; that for a study of primitive and unsocialized feelings the Rorschach may have much potential; and that for the psychotic there is no testing between phantasy and real life. However, the latter conclusion may not be obtainable from the Rorschach data alone (8) unless something is known about the individual's life experience.

Thus, the suggestion put forth here is that structure and unstructure are characteristics of different levels of behavior. That is, because the Rorschach is different as a stimulus situation from other socially-learned-about stimulus situations, it elicits behavior at a different level of organization. Lindner's notion (10) of the correspondence of Rorschach response with the dream fits this view. It would mean that Rorschach data and interpersonal behavioral data are data from different levels of behavioral organization and therefore are not strictly comparable though they may supplement each other in rounding out our understanding of an individual.

This view would explain why Rorschach data predicts social behavior for some and not for others and might resolve the aforementioned dilemma for the "psychotic" record predicting behavior inside a mental hospital but not necessarily outside one. It is suggested that when social behavior occurs at a level of organization divorced from the usual reality considerations we then get a coincidence (isomorphism) of the data of Rorschach behavior with the data of social behavior. It is when this isomorphism pulls apart, that is, when the behavior becomes socialized, (i.e., determined by reality considerations) that we cannot any more make simple inferences of direct relationships. It is in the latter situation that the data from the level of the Rorschach response can be used to supplement our understanding of the total individual by furnishing data about his more "primary" processes.

Such data then can be used to understand an individual through its integration within the context of what we know of his real life space and real actions. It could not be used to predict behavior by simply going from the Rorschach level to the level of social behavior. Schneirla's (18) conceptualization of higher levels of organization as derivative of lower levels, whose organization is however not completely explainable by or reducible to these lower levels is germaine to such an approach as outlined here.

This also applies to Rorschach evaluations of specific attributes. A person might give human movement responses on the test without ever showing capacity for empathetic involvement in real life. True, the test points out a "potential," but what does that mean if an individual's life situation, as well as his perception of his life environment are such that the circumstances for its appearance are nil?

As an example of this, we find that in a level of aspiration experiment (3) in which time estimates of achievement were made in an atmosphere of unreality because true time reference points were absent, a high

relationship between level of aspiration performance and Rorschach indicators of unreality were achieved. However, whether these signs predict the subject's true "unreality" behavior would seem to depend on how closely the level of aspiration test situation approximates real life situations for the particular individual under consideration.

SUMMARY

An attempt is made to draw attention to the unstructured nature of the stimulus situation provided by the Rorschach test, and the meaning that this unstructured condition may have for test interpretation. An experiment is discussed where Rorschach performance predicted behavior for an unstructured situation but not a structured one. Therefore, it is suggested that care must be used before behavior elicited under unstructured situations is utilized to predict behavior occurring within a structured environment. In going from test to real life we may be operating at different levels of behavior organization, and a simple juxtaposition of one to the other would depend on estimating how closely the level of behavior tapped by the test corresponds to the level of behavior in his life space under conditions of more structured and greater reality requirements.

A more thorough self-appraisal here by clinicians may not only lead to further clarification of the nature of the Rorschach test, but also to a more valid understanding of the nature of the test responses and therefore also of their intra-personal meaning.

REFERENCES

1. Fortier, R. H. The response to color and ego functions. *Psychol. Bull.*, 1953, *50*, 41-63.

2. Gibby, R. G., Miller, D. R., and Walker, E. L. The examiner's influence on the Rorschach protocol. *J. consult. Psychol.*, 1953, *17*, 425-428.

3. Gruen, A. The level of aspiration and individual differences. M. A. Thesis. New York University, 1947.

4. Hertzman, M. and Pearce, J. The personal meaning of the human figure in the Rorschach. *Psychiatry*, 1947, *10*, 413-422.

5. Hutt, M. L., Gibby, R., Milton, E. O., and Pottharst, K. The effect of varied experimental "sets" upon Rorschach test performance. *J. proj. Tech.*, 1950, *14*, 181-187.

6. Klopfer, B., Ainsworth, M. D., Klopfer, W. G., Holt, R. R. *Developments in the Rorschach Technique*. World Book Co., New York, 1954.

7. Korner, A. F. Theoretical considerations concerning the scope and limitations of projective techniques. *J. abnorm. soc. Psychol.*, 1950, *45*, 619-627.

8. Kornreich, M. Variations in the consistency of the behavioral meaning of personality test scores. *Genet. Psychol. Monogr.*, 1953, *47*, 73-138.

9. Lazarus, R. S. The influence of color on the protocol of the Rorschach test. *J. abnorm. soc. Psychol.*, 1949, *44*, 506-516.

10. Lindner, R. The clinical uses of content analysis in Rorschach testing. *Psychoanalysis*, 1955, *3*, 12-17.

11. Lord, E. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, *64*, #10.

12. Piotrowski, Z. A. A Rorschach compendium, *Psychiat. Quart.*, 1947, *21*, 79-101.

13. Ranzoni, J. H., Grant, M. Q., and Ives, V. Rorschach "card pull" in a normal adolescent population. *J. proj. Tech.*, 1950, *14*, 107-133.

14. Rorschach, H. *Psychodiagnostics*. New York: Grune & Stratton, Inc., 1942.

15. Sandstrom, C. I. Orientation in the present space. Stockholm, Almqvist & Wiksell, 1951.

16. Schachtel, E. G. On color and affect. *Psychiatry*, 1943, *6*, 393-409.

17. Schachtel, E. G. Projection and its relation to character attitudes and creativity in the kinesthetic responses. *Psychiatry*, 1950, *13*, 69-100.

18. Schneirla, T. C., Levels in the Psychological Capacities of Animals, in: Sellars, R. W., McGill, U. J., & Farber, M. *Philosophy for the future*, New York, Macmillan Co., 1949.

19. Wittenborn, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.*, 1950, *14*, 469-472.

Received February 16, 1957.

The Development of the Rorschach Test in Japan[1]

YASUFUMI KATAGUCHI[2]
National Institute of Mental Health, Japan

In 1925, the Rorschach test was introduced for the first time into Japan by Dr. Y. Uchida, who is known as the author of the 'Uchida-Kraepelin Performance Test." We find a brief introduction to this test described by Dr. Uchida and colleagues in the "Japanese Educational Psychological Research" (Vol. 5, 1930). In this paper they referred to a relation between Rorschach's Erlebnistypus and Kretschmer's typology. Then from 1930 to 1932 Dr. T. Okada (Department of Psychiatry, Kyoto Univ.) wrote five articles in succession on the Rorschach test in the "Neurologia" (Vol. 32, No. 4, No. 5; Vol. 35, No. 2, No. 3). In these articles he described Rorschach's "Psychodiagnostik" in detail and suggested his own scoring system.

But after his, no research of interest to clinical practitioners was published. This might be because there were limitations in the clinical effectiveness of the Rorschach test so long as it adhered to Rorschach's original interpretation. Of course Rorschach's gifted intuition was very excellent, but use of his method had come to a standstill. This test was born in Switzerland, but we may say that it was after all developed in America, especially by Drs. S. J. Beck and B. Klopfer, and brought by them to its present-day prosperity.

When the Rorschach test as developed in America was introduced, psychologists in Japan became interested in it. Prof. H. Motoaki (Department of Psychology, Waseda Univ.) was the first to generalize this test. About 1941 he was still examining the 'Psychodiagnostik," but later he adopted Beck's system and further took in Klopfer's movement categories. In 1952 he published his "Manual of Clinical Diagnostic Tests of Personality," a most popular Rorschach text book in Japan. He also devised a modified series of inkblots that consisted of eight cards: the "Waseda Modified Rorschach."

Dr. R. Satake (Kanazawa Classification Office of Juvenile Delinquency) was the first to introduce the group-method of this test in this country. Now he is trying the Sodium Amytal Rorschach (SAR) to make clinical diagnosis more effective. Meanwhile, Dr. G. Nagasaka (Department of Psychiatry, Osaka Univ.) began to study this test intensively from 1945. At first he followed Klopfer's system, but afterward developed his own original ideas and constructed his so-called "Dynamic Tabulation Method." This method is a new approach to sequence analysis and is useful in the clinical field. Now he is engaged in systematic research with many collaborators. This group has already produced forty papers since 1947.

Prof. H. Kodama (Child Research Institute, Japan Woman's Univ.) became interested in this test after World War II. From the beginning he planned large-scale normative studies and seven years have already passed since he started. The definitive norms of the Rorschach responses in Japan are expected from his researches. He has used more than 4,000 subjects from three to eight years old who represent various socio-economic levels.

Recently, Mr. E. Murakami

[1] The author would like to express his gratitude to Dr. B. R. Forer who encouraged him to write this report. He also wishes to thank Prof. Y. Uchida, Prof. H. Motoaki, Dr. G. Nagasaka, Prof. Kodama, and Dr. M. Kato for their helpful suggestions.

[2] Clinical psychologist, research worker in the National Institute of Mental Health, Ichikawa city, Chiba, Japan.

(Department of Psychiatry, Nagoya Univ.) influenced by Dr. J. DeVos, who visited this department several years ago, took interest in content analysis and is now working on it. More recently, Prof. T. Imura (Department of Psychiatry, Nihon Univ.) and his pupils have begun studies on clinical applications of the Z-Test.

Thus, research on the Rorschach test in Japan has prospered year by year. So the clinical psychologists and the psychiatrists of the psychiatric department of hospitals, the Classification Office of Juvenile Delinquency, Child Guidance Clinic, and Mental Hygiene Clinic have come to be more or less interested in this test. In Japan the most favorite of the projective methods are the Rorschach and TAT, but the SCT, Szondi, Picture Frustration Test and Bender-Gestalt are also used.

The numbers of the papers (on the Rorschach) read in the annual meetings of the representative learned societies in Japan are as follows:

	Japanese Psychological Association	Japanese Association of Psychiatry and Neurology
1950	1	1
1951	3	2
1952	8	1
1953	5	4
1954	11	2
1955	12	3
1956	23	6
1957	1	5

Text books or introductory manuals of the Rorschach test in Japan are as follows:

H. Motoaki: *Manual of clinical diagnostic tests of personality*. Tokyo: Kaneko Book Co., 1952.

H. Kodama: A normative study of Rorschach responses in Japanese. *Psychology Series, 7*, II. Tokyo: Nakayama Book Co., 1953.

R. Satake: *The Rorschach test. Abnormal Psychology Series*, Part 1, B2. Tokyo: Misuzu Book Co., 1954.

Y. Kataguchi: *The psychodiagnostic technique (Shinri-shindan-ho). The Rorschach technique*. Tokyo: Maki Book Co., 1956. (This book was reviewed by Dr. B. R. Forer in this journal, 1956, *20*, No. 4.)

H. Motoaki: *Manual of group method of the Rorschach test*. Tokyo: Kaneko Book Co., 1957.

The bibliography of Rorschach articles published in Japanese journals is as follows:

Psychiatria et Neurologia Japonica

T. Okada: Experimentelle Studie über die Rorschach'sche "Psychodiagnostik." 1930, *32*, No. 5, 43-55.

T. Okada: Über die Einteilung der Reaktionen und Analyse der abnormen Aussage beim Rorschachschen Formdeutversuch in der Psychodiagnostik. 1932, *35*, No. 2, 39-82.

G. Nakasaka: Studies on Rorschach test. 1952, *54*, No. 4, 219-253 (English abstract).

T. Nishigori: Die psychiatrische Erforschung der Strafgefangenen durch den Rorschachschen psychodiagnostischen Versuch. 1953, *55*, No. 1, 79-90 (German abstract).

K. Takahashi: A study of the Group Rorschach test. 1954, *55*, No. 9, 22-39 (English abstract).

M. Kuribayashi: An investigation of the Rorschach test for the Wake-amine poisoning. 1955, *57*, No. 7, 1-8 (English abstract).

T. Hamanaka: A psychodynamic study on the mechanism of development of psychogenetic tic. 1956, *58*, No. 9, 1-23 (English abstract).

Japanese J. Psychology

H. Motoaki: A study of "good form" responses in the Rorschach test. 1942, *17*, No. 2, 273-282.

Y. Niimi et al: Galvanic skin responses during Rorschach test administration. 1956, *27*, No. 3, 175-184 (English abstract).

E. Murakami and Y. Fmi: A study on the Rorschach ranking test. 1956, *27*, No. 3, 185-192 (English abstract).

K. Murakami and H. Noda: A study on diagnosis of intelligence by the Rorschach test. 1957, *27*, No. 5, 375-377.

Japanese J. Educational Psychology

S. Tsuji, K. Miki et al: Some investigations on the significance of children's ranking of their inmates by preference. 1954, *2*, No. 1, 18-29 (English abstract).

K. Ohira: A study on the characters of twins by Rorschach test. 1955, *3*, No. 1, 11-17 (English abstract).

T. Ikeda and H. Kawai: A pattern analytic approach of the Rorschach technique, applied to institutional children. 1957, *4*, No. 4, 15-19 (English abstract).

J. Mental Health

Y. Kataguchi and H. Dendo: A study of the personality of the traumatic neurosis by Rorschach test. 1955, No. 4, 30-41 (English abstract).

M. Saji and Y. Kataguchi: A study on the measurement of the effect of psychotherapy. 1956, No. 4, 65-112 (English abstract).

Received June 10, 1957

Alleged Rorschach Anxiety Indices in Children

EUGENE E. LEVITT
Illinois Institute for Juvenile Research[1]

Based on a review of the literature, Eichler (3) found that there are some 15 Rorschach variables which are commonly regarded by projective test experts as anxiety indicators. It is not suggested that these indices necessarily hold for child populations, or that we should expect them to. On the contrary, it has been pointed out by Halpern (4) and Paulsen (5) that pathological signs need not be common to adults and children because of the development aspect of the child's responses. Nevertheless, we have not yet ascertained experimentally that such commonality does *not* exist. The purpose of this paper is to present experimental data bearing on the point. These data consist of (a) comparisons between samples of normal children and clinic patients on nine of the variables listed by Eichler (3), and on seven other factors related to them, and (b) the relationships between variables and scores on the Children's Manifest Anxiety Scale (CMAS) for the clinic sample. The CMAS is a 42-item true-false verbal inventory patterned after the Taylor Anxiety Scale. Its development has been described elsewhere (2). In the clinic administration, each item was read by the psychologist to the child, who indicated his responses on a response sheet. The CMAS has been found to differentiate normal and disturbed children and to be related to psychiatrists' ratings of severity of symptoms for disturbed children (6).

SAMPLE AND RESULTS

The sample of disturbed children consisted of 39 unselected cases of the age range 8.08—13.25 years, who were seen diagnostically at a community guidance clinic in Chicago. The mean age was 10.70 years with an *SD* of 1.42. There were thirty boys and nine girls, the usual ratio at this clinic. The average IQ based on 34 cases for whom scores were available was 103.62 with an *SD* of 16.61. The mean score on the CMAS for all 39 cases was 20.49 with an *SD* of 7.80, as contrasted with a mean of 13.23 for 322 Chicago school children tested by Rynerson and Levitt (6) and a mean of 17.13 for 386 Iowa school children examined by Castaneda, McCandless and Palermo (2).

The data for the normal sample were taken from the report of Thetford, Molish and Beck (8). The 155 children furnishing these data were all from Chicago public schools, and were described as "free from overt behavior problems discernible to their teachers," of "average academic achievement," and "within normal range" on one of three standard intelligence tests. The age range was 6 to 17.5 years with a mean of 10.75 and an *SD* of 3.17. There were 83 boys and 72 girls in the sample.

The samples are satisfactorily equated for mean age, and very probably for intelligence. The variabilities of age are different, as are the sex ratios. The possible effects of these differences on results will be discussed later.

Nine of the indices mentioned by Eichler (3) are reported for the normal sample. These are: R, W, P, Hd, M, Dd, A%, Sum Y, and Sum C. Data for the related factors, W%, Y, FY, YF, C, FC, and CF, are also given. Except for R, W, W%, A% and P, distributions tended to be skewed for one or both samples. Differences between groups on these five variables were analyzed by *t*-tests.

[1] Now at the Indiana University Medical Center.

TABLE I—Comparison of Normal and Disturbed Children on Symmetrically Distributed Variables

Variable	Sample Mean Normal	Sample Mean Clinic	*t*	Variance Ratio
R	27.15	25.82	0.58	1.26
W	3.14	7.26	5.24**	1.99*
W%	12.55	30.77	6.07**	2.36**
A%	47.95	42.03	2.15*	1.16
P	4.54	5.00	1.67	1.91*

* Significant beyond the .05 level.
** Significant beyond the .01 level.

TABLE II—Comparison of Normal and Disturbed Children on Asymmetrically Distributed Variables

Variable	Dichotomized at	Sample Scoring Higher	Chi-square	*P*
Hd	median	normal	4.12	.05
Dd	2.5	normal	3.74	.054
M	median	clinic	21.75	<.01
Sum C	median	clinic	4.32	.04
Sum Y	median	clinic	4.30	.04
C	0.5	normal	0.66	ns
FC	0.5	....	0.00	ns
CF	0.5	clinic	10.45	<.01
Y	0.5	normal	2.99	ns
FY	0.5	....	0.00	ns
YF	0.5	clinic	16.09	<.01

Variance ratios were also computed. The skewed distributions were artificially dichotomized and chi-squares were computed. The results for the five symmetrically distributed variables are shown in Table I.

The results of the analyses of the eleven asymmetrically distributed variables are shown in Table II.

Nine of the 16 analyses of average frequency or percentage are significant at the .05 level or beyond. The variables are W, W%, A%, Hd, M, Sum C, Sum Y, CF and YF. The chi-square for Dd falls short of the .05 level by only .004. The clinic sample is higher on all of these factors except A%, Hd and Dd. Variance ratios for W, W% and P also reach at least the .05 level. The normal group shows more variability in P and the clinic sample in the others.

The correlations between most of the Rorschach indices listed by Eichler and scores on the CMAS for the clinic sample are given in Table III. Relationships were analyzed using the Pearson *r* for ten variables which were symmetrically distributed. Three factors were analyzed by dichotomizing the distribution of CMAS scores based on a logical split of the Rorschach variable, and computing *t*-tests. One analysis is a *F*-ratio based on a trichotomous split.

None of the analyses of Table III is significant, or even approaches significance. In fact, no *P* value is as low as .20.

Discussion and Conclusions

The consensus holds that anxiety indicators are low scorings on R, W, P, M and Sum C, and high scorings on Hd, Dd, A and Sum Y. The results in Tables I and II are almost unanimously opposite to the predicted direction. The clinic group scores higher on W, M and Sum C, and lower on Hd, Dd, and A%. There is no difference in R. Only Sum Y, probably the most widely accepted anxiety sign, is in the predicted direction.

Since the samples differ with re-

TABLE III—Relationships Between Rorschach Indices and the Children's Manifest Anxiety Scale for the Clinic Sample ($N = 39$)

Symmetrically Distributed Variables	r
R	.14
W	.11
W%	.02
M	.11
Sum C	.14
P	—.08
A%	.16
F	.11
A	.16
M%	.01
P%	—.09
Asymmetrically Distributed Variables	t
Hd	0.88
Sum Y	0.12
Dd	1.49*
Ad	0.18
An	0.17

* An F-ratio based on a trichotomous breakdown of this variable.

spect to proportions of boys and girls, and in variability of age, it is possible that these circumstances have affected the results. Data as a function of age are reported by Thetford et al. (8). No age differences in the normal sample were found for Y, YF, C, FC, W%, A%, Hd and Dd. Frequencies were roughly positively correlated with age for R, P, and FY, so that the net effect of the absence of the highest and lowest age groups in the clinic sample would be nil. Frequencies tend to be peaked at the higher age groups with no other differences for W, M, CF, Sum C and Sum Y. Hence we would expect that the lack of children older than 13 years in the clinic sample would result in reduced mean frequencies for these variables. Since the clinic group actually exceeds the normal for all of the variables, we must conclude that age was not a factor influencing results appreciably.

Unfortunately, Thetford et al. (8) do not present comparative data by sexes. Accounts of sex differences in the literature such as those of Ames et al. (1) and Stavrianos (7) are somewhat in conflict. There seems to be general agreement that boys score higher on R and W%. This is also true of the clinic sample, although the numbers of cases of each sex are too small for reliable statistical analysis. The Ames et al. data are derived from a high IQ group[2] which is not comparable to the samples tested here. However, in the absence of more appropriate data, we will use the Ames study as a guide in determining sex differences in normal samples. Then, by algebraically summing the expectation in the Thetford sample with the sex differences found in the clinic sample, we find that the net will explain differences in W%, and CF. Differences or lack of differences in M, Sum C, Y, A%, Dd, and R are contrary to expectation. In general, the conclusion is that the sex factor does not appear to have markedly affected the comparison.

Even if we hypothesize the operation of some uncontrolled variable affecting results, it is certainly unlikely that it could be momentous enough to turn significantly larger averages of the clinic group into significantly larger averages for the normals. At best there would be no differences. Differences in variability such as were found for W, W%, and P are usually difficult to interpret unless they are encompassed by some theory.

In general, the conclusion must be that none of the alleged anxiety indicators studied here holds up with the exception of Sum Y, the total of the unweighted shading responses. This conclusion is partly supported by the data of Table III which indicate that none of the alleged indices is related to an independent measure of anxiety in the clinic sample. When viewed alone, however, the absence of relationship could not be considered as conclusive. The CMAS is designed

[2] The mean IQ is not given, but the frequency distribution indicates that it is at least 120.

to measure manifest anxiety, while the Rorschach is presumed to tap a deeper level. The lack of correlation is therefore not completely surprising.

The possibility that many of the conventional adult anxiety indices are reversed in children may sound outlandish, but should not be discarded hastily. The results of this study suggest that such a notion ought to be afforded further experimental investigation.

Summary

A group of child clinic patients was compared with a group of normal children on a series of Rorschach variables commonly regarded as anxiety indices in adults. Of nine significant differences, only the frequency of shading responses was in the predicted direction. Within the clinic group, no variable was found to be related to scores on the Children's Manifest Anxiety Scale. Possible effects of differential sex ratios and variabilities of age between the two samples were discussed, and it was determined that these did not influence results significantly. It was concluded that as far as the present data go, only frequency of shading responses can be considered to be an anxiety indicator in children.

References

1. Ames, Louise B. et al. *Child Rorschach responses.* New York: Hoeber, 1952.

2. Castaneda, A., McCandless, B. R., and Palermo, D. S. The children's form of the Manifest Anxiety Scale. *Child Develpm.,* 1956, *27,* 317-326.

3. Eichler, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.,* 1951, *46,* 344-355.

4. Halpern, Florence. *A clinical approach to children's Rorschachs.* New York: Grune & Stratton, 1953.

5. Paulsen, Alma A. Personality development in the middle years of childhood: a ten-year longitudinal study of thirty public school children by means of Rorschach test and social histories. *Amer. J. Orthopsychiat.,* 1954, *24,* 336-350.

6. Rynerson, Mary N., and Levitt, E. E. A comparison of the performances of normal and disturbed children on the Children's Manifest Anxiety Scale. Paper read at the meeting of the Society for Research in Child Development, 1957.

7. Stavrianos, Bertha. An investigation of sex differences in children as revealed by the Rorschach method. *Rorschach Res. Exch.,* 1942, *6,* 168-175.

8. Thetford, W. N., Molish, H. B., and Beck, S. J. Developmental aspects of personality structure in normal children. *J. proj. Tech.,* 1951, *15,* 58-78.

Received March 3, 1957

Quantification of the Bender-Gestalt Recall: A Pilot Study[1]

TOM D. OLIN AND MARVIN REZNIKOFF
The Institute of Living

INTRODUCTION

While the Bender-Gestalt test was originally devised to investigate disturbances in visual-motor performance resulting from organic brain disease, it has been found to be useful in the study of a variety of personality disturbances of a predominantly functional nature. With new applications of the test, modifications in the standard administration technique have been introduced. Of these, the recall procedure has provoked the most interest and is now frequently used to supplement the material obtained from the standard procedure. However, in rather marked contrast to the effort devoted to quantification of the original copying procedure (3, 6, 7, 10, 11), comparatively few studies are available which attempt to quantify the recall in a similar fashion. In addition, even the procedure used in obtaining the recall has been subject to wide variation. As Reznikoff and Olin (12) have pointed out, this failure to standardize the recall administration makes inter-study comparisons difficult and of doubtful validity.

A survey of the literature relating generally to the recall procedure reveals that the major interest has been in comparing the number of designs recalled by one group with the recall of another. For the most part, interest has been focussed on the qualitative aspects of the reproductions only to the extent of establishing criteria for scoring a design as remembered. Moreover, the criteria utilized varies considerably from one investigator to another. For example, Bossom (4) scores one point for a "correctly completed" reproduction and zero for all others; Tolor (17) counts both "whole and part figures" in his measurement of the recall; and Gobetz (6), along with a number of other investigators, provides no published criteria for measuring the recall. However, despite the differences currently existing both in the measurement and administration of the recall, a number of relatively consistent findings emerge from the available studies. These findings, in a general way, suggest that organic patients are poorest in recall with schizophrenics, neurotics and normals coming next in this order. With the aim of improving the diagnostic usefulness of the recall procedure, the purpose of the present research was to investigate a more explicit and detailed method of quantifying the essentially qualitative recall data. It was felt that a scoring system utilized for the recall productions should be equally applicable to the copied designs so that direct comparison could be made between these two phases of the test.

Of the published investigations dealing with the recall, four are specifically concerned with quantification. Arazi (2) lists eight deviations occurring in the recall, five of which she feels are comparatively minor distortions and are not used in her group comparisons. Employing the remaining three indices, involving major distortions of the original stimuli, she was able to obtain significant differences between schizophrenics and normals. These three indicators, however, are only a very small portion of the deviations commonly found in recall productions. Moreover, they rarely and sometimes never occur in the copied productions, thus limiting their usefulness for copy-recall comparisons.

[1] The authors wish to thank Miss Louise Hewson, Director, Department of Psychology, Neurological Institute, New York, for her assistance in obtaining the organic cases used in this research.

Shapiro, et al. (14), using a modified and reduced series of Bender-Gestalt designs in their geriatric studies, developed a scoring method consisting of five levels of accuracy for each design reproduction whether occurring on the standard copy phase or on the recall. They found that organic subjects obtained poorer scores on both the copy and recall than a control group of non-organic subjects. While this investigation demonstrated rather clearly the usefulness of recall quantification, and provided adequate criteria for scoring protocols, the difference in stimuli together with the advanced age of his population prevent any direct transference of his system to the standard stimuli and younger age groups.

Gobetz (6) developed an extended series of indices most of which was applicable to both the standard and recall procedures. In comparison of neurotics and normals, however, he found no difference in recall which held up under cross validation. While his scoring method contains a number of unique features which appear to be potentially of use in evaluating both the recall and the copy phase, little data is presented which might serve to clarify the value of his specific indicators with other groups of subjects, e.g., organics and schizophrenics. Furthermore, a number of his indices are not applicable to both the recall and the copy phases, and thus cannot be used for such comparisons.

The system devised by Peek and Quast (11) is also designed to provide a means of scoring both the copy and recall phases using one basic series of indices. Their method includes a number of deviations not found in the above scoring procedures in addition to several indicators which are predominantly encountered in the recall phase alone. Since these investigators have thus far presented no data using their system with either the recall or copy procedures, the value of their method remains uncertain, despite the fact that many of the indices appear to have considerable face validity.

Though not specifically concerned with quantification of the recall, Pascal and Suttell (10) have developed a scoring system for use with the copy phase which has been utilized in a large number of studies. The results of these investigations indicate that it is capable of demonstrating comparatively consistent group differences when the copy protocols are scored according to their detailed criteria. Parker (8) found that brain-injured subjects scored higher, indicating worse performance, than did non brain-injured subjects; Robinson (13) found that paretics scored higher than did schizophrenics; Addington (1), Sonder (15), and Suttell and Pascal (16) have found that schizophrenics score higher than do normal controls; Bowland and Deabler (5), using a group administration of the Bender-Gestalt, found that organics scored highest with schizophrenics, neurotics and normals coming next, in this order of deviation from the original stimuli. Other investigations using the Pascal and Suttell system tend to concur on the nature and direction of these group differences.

Considering the demonstrated ability of the Pascal and Suttell procedure (hereafter referred to as the P & S system) to differentiate a variety of groups of subjects, it was felt that this system might provide a valuable basis for developing a comparable scoring system for the recall; a system which, hopefully, could then be used as a companion method with the procedure now used solely for the copy phase. Informal analysis of a large number of recall protocols revealed that the specific indices included in the P & S system were for the most part scorable for the recall. In addition, the recall was found to contain an insignificant number of deviations which could not readily be included in the P & S system.

Briefly, the P & S system consists of

105 scorable deviations of which seven deal with general factors related to the placement of the designs on the page, while 98 are concerned with specific deviations from the standard reproductions of Designs 1 through 8. Each item is weighted in accordance with its relative frequency in the standardization population with the summed weights of the 105 items constituting the 'Raw Score' for the particular protocol. It is, of course, possible to obtain total scores for each of the designs considered independently in the same manner. For subjects having at least one year of high school and within the age range of 15 to 50 the Raw Score may be converted into a standard, or 'z-score.' The mean of the z-score distribution is 50 and the standard deviation is 10. This permits ready comparison of the individual protocol with the non-patient standardization group with regard to the degree of drawing deviation shown. A z-score of 72 or over suggests, according to the authors, that the subject is "in need of psychiatric help." Extensive scoring criteria are provided in the authors' text together with a number of practice protocols; this makes it a relatively simple system to learn to use competently. Only Peek and Quast have similar, but far fewer examples to illustrate their scoring criteria. The presence of these specific illustrative criteria in the P & S system is an additional reason for its selection as a procedure for recall quantification.

USE OF THE P & S SYSTEM IN EVALUATING THE RECALL

A number of theoretical and practical problems arise when the attempt is made to use a system devised and standardized on the copy phase with the recall phase. Possibly the major consideration involves the justification of transferring deviation weightings, ranging from '1' to '8', since these weights were derived from a comparison of patient and non-patient frequencies on the copy phase and not on the recall. On the surface, it is quite possible that specific deviations occurring on the recall would occur with differing frequencies and, therefore, should be weighted differently in computing a recall score. Scoring on the basis of the copy weights might, thus, obscure real differences which exist between groups rather than clarifying them.

Another problem concerns the fact that certain deviations found in the recall productions were not found by P & S in their analysis of the copy and thus no provision was made for their scoring. Examples of this would be the reversal of slant in the eleven rows of Design 2, the reversal of apex in the vertical rows of Design 3, and the closure of the three sided square of Design 4. Analysis of a large number of recall protocols, however, revealed that deviations of this sort occurred very infrequently. Because of this, and since the inclusion of deviations of this sort would obscure copy-recall comparisons, it was decided to omit them from the present recall quantification.

The problem of contaminated and additional productions is related to the above difficulties in transferring the P & S copy system to the recall, since this type of deviation does not occur in the copy phase. By 'contaminated' responses is meant a fusion or combining of two or more designs, or design parts, into one final production. The most commonly found contamination is the joining of a portion of Design A with a portion of Design 4, resulting in a distorted reproduction resembling both A and 4. While this particular deviation is not specifically encountered in the copy, it was decided that it was closely related to the P & S scoring category of 'distortion,' and that it could be scored under this category without unduly violating this concept. In the above example of contamination, the problem of whether to score the production as a distortion of Design A or Design 4 was handled by scoring

'distortion' for the design which it most resembled as determined by the position and relative prominence of the elements. In practice, this judgment was made without significant ambiguity. With regard to the 'recall' of designs which bear no apparent resemblance to any of the original stimuli, the so called 'additional' productions, which occur very infrequently, the decision was made to eliminate them from any scoring consideration.

One additional problem arises in so using the P & S scoring system, namely, Design A was not quantified by the authors. The reasons for this remain obscure since P & S make no mention of the basis for omission of this design in either the major text, or in a previous publication dealing with the system by Pascal (9). Because all other attempts at quantification of the copy phase have scored Design A, and since all efforts at measuring the recall phase have considered Design A as an integral part of the total protocol, it was deemed advisable to score this design along with the eight others. To do this it might have been possible to utilize available scoring systems obtained from other investigators. This was felt to be undesirable, however, as it would introduce scoring categories which would be quite divergent from those found in the P & S scoring of the other eight designs.

Analysis of the 98 specific scored deviations reveals a comparatively large overlap of scoring categories from design to design. For example, the category 'Second Attempt' is scored in the same manner for each of the eight designs. Of the total of 98 items, only 39 different categories are used, and of these, only 19 are specific to a single design, while three are encountered on each design. Moreover, designs containing common elements, dots for example, have certain scoring categories in common, referring to the elements which the several designs have in common. Considering this extensive overlap of scoring categories among the various designs, it was felt that it would be consistent with the P & S system to develop a series of indices for Design A reflecting the common elements already existing within the system. Since Design A has most in common with Design 4 with respect to the nature of the elements and their interrelationship, Design 4 was chosen as the basic model from which to develop a scoring system for Design A. In addition to criteria and specific items obtained from Design 4, however, several indices in the final series came from Designs 7 and 8, and one was suggested by the 'Configuration' series of indices.

A total of 19 indices were finally developed for the scoring of Design A. These are listed below together with their respective weights; that is, those weights which were given to similar deviations when they occurred on the other designs from which the particular Design A item was taken. Perusal of a number of recall protocols suggested that these 19 indices were sufficiently extensive to handle all of the deviations found in the recall of Design A. In the interests of brevity, criteria for scoring each of the indices will be given only for those items in which a significant modification of the P & S criteria has been made to relate it specifically to Design A. The notation '4, 2' following the number in parenthesis, which is the weighting of the deviation, refers to that scoring item explained in the P & S manual under Design 4, Item 2, and so forth. The scoring indices for Design A are as follows:

1. *Asymetrical circle* (3)
   a. One diameter is 2/3rds or less the length of the other; measurement to be taken at the longest and shortest diameters of the circle respectively.
   b. One half of the circle is reproduced in a manner significantly different from the other half.
   c. Angles are present in the circle.
2. *Asymetrical square* (3)
   a. One angle in the square is approx-

imately twice that of another.
   b. One of the lines is 2/3rds or less of another.
   c. One half of the square is reproduced in a manner significantly different from the other half.
3. *Break in circle* (4) '4,2'
4. *Ends not joined* (8)
   a. When two or more gaps in the square occur of 1/16th inch or more.
   b. When one gap occurs which is 1/8th inch or more; if this occurs at the juncture of square and circle, an additional gap of 1/16th inch is needed.
5. *Square not centered* (1) '4,3'
6. *Square not joined to circle* (8) '4,5'
7. *Square overlapping* (8) '4,5'
8. *Angles extra in square* (3) '7,2'
9. *Extra scattered dots or dashes* (3) '7,4'
10. *Double lines* (1 ea.) '7,5'
11. *Guide lines* (2) '4,10'
12. *Tremor* (4) '4,8'
13. *Workover* (2) '8,9'
14. *Square rotation* (3)
    The square is rotated approximately 45° thereby appearing to 'rest' on one side.
15. *Second attempt* (3 ea.) '4,11'
16. *Rotation* (8)
    The square is rotated in such a manner that the resulting configuration of square and circle has been rotated 45° or more.
17. *Distortion* (8)
    a. When the design has been contaminated with another.
    b. When there is other marked distortion of the original stimulus as, for example, a triangle substituted for the square.
18. *Design missing* (8)
    a. When one of the two figures is missing.
    b. When at least 1/3rd of either the circle or the square is missing.
19. *Relative size* (8)
    a. When the largest diameter of the square is twice the length of the largest diameter of the circle, and vice versa.
    b. When the apparent area of one part is approximately 1/4 the area of the other.

While the use of weights for each of the nineteen indices is, of course, largely arbitrary, it was felt that they corresponded closely to the weights given comparable deviations on the other designs. In addition, it may be noted that in initially determining the weights applied to the copy phase indices, P & S gave the same weighting to the same item on whichever design it was encountered. Thus the item 'Rotation,' which is scored for all eight designs in the P & S system, was given a weighting of '8', whereas the proportion by which patients differed from non-patients in the standardization population varied considerably depending upon the particular design in which rotation was scored. These features of the original weighting procedure were felt to mitigate in some measure the apparent arbitrary character of the above assignment of weights to Design A. As with the other designs, the recall score for Design A is obtained by totaling the weights of all indices found to be present in the reproduction of the design. The higher the resulting score the less accurately has the design been recalled.

SUBJECTS AND PROCEDURE

Two of the subject groups used in this investigation were described in a previous study reported by the authors (12) dealing with the number of designs recalled. These two groups consisted of 33 organic patients from the Neurological Institute of New York, and 50 schizophrenic patients hospitalized at the Institute of Living. No organic patient had a concommitant functional diagnosis and all of the schizophrenics were free of any organic pathology. The patients ranged between 18 and 49 years of age and had obtained full scale Wechsler IQ's ranging from 80 to 128. While the mean age of the schizophrenic group was significantly above that of the organic group, ages 37 and 31 respectively, the correlations between age and the number of figures recalled was not significant. The schizophrenics had a somewhat higher mean IQ than did the organics, 107.5

and 100.2 respectively, however this difference was not statistically significant.

In order to obtain normative data related to recall quantification, a third group of subjects was included in the present study. This consisted of 33 student nurses who were in the final phases of training at the Institute of Living. All of the subjects but one were between 19 and 23 years of age. While no measures of intellectual functioning were available for this group, it was assumed that the group as a whole was of average to bright normal intelligence. Because the group of nurses differed widely from the other two groups in such factors as sex composition, occupation, age, and possibly IQ, it was felt that this group was inadequate as a rigorous control for the patient groups. The comparison data to be presented, therefore, must be considered as merely suggesting trends which must be verified with other more suitable control populations before definite conclusions can be drawn.

The two patient groups were given the Bender-Gestalt copy and immediate recall procedures as a routine part of their psychological evaluations. The nurses were given the same tests on a voluntary basis at the close of their training. All of the copy and recall productions were scored by the senior author using the P & S scoring system for the copy phase and the present adaptation of it for the recall. In addition, the copied production of Design A was also scored according to the procedure outlined above.

Results

Reliability

In order to determine the reliability of the scoring procedures the records of the two patient groups were re-scored after several weeks had elapsed. The correlations for combined patient groups between the original scoring and the re-scoring for the copy and recall phases when the P & S indices for Designs 1 through 8 were totaled were .93 and .84 respectively. When Design A was considered separately, the respective correlations were .81 for the copy phase and .87 for the recall.

Accuracy of Recall of Organic, Schizophrenic and Normal Groups

Table I presents the mean qualitative scores of the organic and schizophrenic patients for each of the nine designs treated individually. The *t* test was used to determine the significance of the mean differences. As can be seen, the organics' reproductions were qualitatively inferior to the schizophrenics' on all nine designs. The differences between the groups were significant at the .10 level for Designs A, 3, 5 and 8, and at the .05 level for Design 4.

Table II presents similar compari-

Table I. A Comparison of the Recall Scores Obtained by the Organics and Schizophrenics for Each of the Bender Designs

Design	Percent Recalling Design		Mean Recall Score*			
	Organics	Schizophrenics	Organics	Schizophrenics	t	P
A	76	60	9.16	6.06	1.76	.10
1	58	78	4.11	3.82	.31	.80
2	36	70	6.67	5.17	.68	.50
3	27	26	9.56	5.31	1.93	.10
4	18	20	16.67	9.60	2.14	.05
5	70	58	8.00	5.27	1.68	.10
6	76	74	7.88	6.57	.93	.40
7	45	58	16.69	15.10	.53	.60
8	85	76	6.79	4.79	1.71	.10

* Each of the nine designs has a different maximum score.

TABLE II. A Comparison of the Recall Scores Obtained by the Organics and Normals for Each of the Bender Designs

Design	Percent Recalling Design		Mean Recall Score*			
	Organics	Normals	Organics	Normals	t	P
A	76	88	9.16	6.08	1.82	.10
1	58	85	4.11	2.92	1.15	.30
2	36	76	6.67	3.96	2.19	.05
3	27	36	9.56	7.75	.96	.40
4	18	21	16.67	14.14	.69	.50
5	70	79	8.00	4.58	2.51	.02
6	76	94	7.88	4.35	2.80	.01
7	45	85	16.69	11.21	2.16	.05
8	85	88	6.79	3.03	3.11	.01

* Each of the nine designs has a different maximum score.

TABLE III. A Comparison of the Recall Scores Obtained by Schizophrenics and Normals for Each of the Bender Designs

Design	Percent Recalling Design		Mean Recall Score*			
	Schizophrenics	Normals	Schizophrenics	Normals	t	P
A	60	88	6.06	6.08	.01	.99
1	78	85	3.82	2.92	1.05	.30
2	70	76	5.17	3.96	.83	.50
3	26	36	5.31	7.75	1.56	.20
4	20	21	9.60	14.14	1.53	.20
5	58	79	5.27	4.58	.52	.70
6	74	94	6.57	4.35	2.04	.05
7	58	88	15.10	11.21	1.80	.10
8	86	85	4.79	3.03	1.70	.10

* Each of the nine designs has a different maximum score.

son data for the organics and the normals. It can be noted that the organics' ability to reproduce the designs accurately from memory was even more inferior when contrasted with the normal group. Design A reaches the .10 level of significance, whereas Designs 2, 5, 6, 7, and 8 reach the .05 level or better. Of the nine designs the superiority of both the schizophrenics and normals over the organics was most clearly demonstrated on Designs A, 5 and 8.

Table III reveals that when the schizophrenics are compared with the normals, the differences in quality of recall are not nearly as marked or so consistently in one direction. The schizophrenics were able to reproduce Designs 3 and 4 with greater accuracy than did the normals, although the .20 level of confidence obtained indicates that this difference is insignificant. Design A was reproduced with the same accuracy for both groups. In all, differences between the two groups at or beyond the .10 confidence level occurred on only three designs, 6, 7, and 8, with only Design 6 reaching the .05 level. In each of these three instances the normals surpassed the schizophrenics in the accuracy of their reproductions.

Indicated also in Tables I, II and III are the percentages of the three groups recalling each of the nine figures. An examination of these percentages shows that Designs 3 and 4 are clearly the most difficult figures for members of all three groups to recall while Designs 6 and 8 are among the easiest. These findings are consistent with results obtained by other investigators (4, 17). The mean number of designs recalled was 4.88 for the organics, 5.30 for the schizophrenics, and 6.52 for the normals. In this tabulation a design was consid-

ered to have been recalled if it could be scored according to the modified P & S system regardless of the accuracy with which it was reproduced. Again, these findings are similar to those obtained in a number of other studies previously cited.

Since it was felt the practice of weighting the various distortions in accordance with the P & S system, and the combining them to obtain a single score for each of the nine designs might possibly obscure some of the differences, it was decided to compare the number of times each of the total of 117 indicators was scored in the three groups. The figure 117 includes the original 98 indices for Designs 1 through 8 together with the 19 additional items for Design A alone. In addition such an analysis might reveal particular deviations in drawing which would be useful to the clinician in the absence of more formal scoring.

While this individual item analysis could be handled by a straight forward tabulation of the number of subjects in the three groups for whom a particular deviation was scored, with certain of the specific items such a procedure might lead to somewhat misleading results. This is due to the fact that a subject can receive more than one scoring for certain of the indices. For example, one of the schizophrenic subjects perseverated the number of dots on Design 1 on the copy phase, and was scored thirteen times for Item 5 of Design 1, 'Number of dots.' Another subject, an organic, was scored 6 times for the presence of 'Double line' on Design 7, Item 5 on the recall phase. Out of the total of 117 separate items, 19 were found which might be scored for a particular subject more than once; the remainder, or 98, were scorable only once. Of the 19 items, only six were found to have been scored more than once with sufficient frequency to warrant a different analysis. These six items were: 'Double line' occurring on Designs A, 6, 7 and 8; 'Number of dots' on Design 1; and 'Number of columns' on Design 2. To bring these items in line with the others in which the maximum frequency of scoring per subject was one, it was decided to score each of the six deviations as either present in the individual protocol or not present, regardless of the number of times the item occurred for the particular subject. In addition, 'Double line,' occurring on Designs A, 7 and 8 was scored separately if a double line was encountered (a) one or more times, and (b) two or more times, since examination of the obtained frequencies suggested that certain group differences might be clarified in this manner.

Table IV reports those indicators which produced differences between the groups significant at or above the .10 level of probability when the three groups were considered two at a time. The number of organic cases available had increased from 33 to 37 at the time these computations were done, and accordingly the four additional cases were included in the organic group.

Of the twenty indicators reaching the .10 level of confidence or better, ten differentiated between the organic and schizophrenic groups, and six between the organic and normal groups. Of these 16 indicators, the organics were superior in only two, the presence of 'Guide lines' and 'Second attempt' on Design 7, when compared with the schizophrenic group. Of the remaining four indicators, which discriminated between the schizophrenics and normals, two were found a greater number of times in the normal group, while two occurred considerably more often among the schizophrenic patients. It is recognized that all of the differences reported in Table IV are within chance expectation for the number of comparisons made, 117 for each pair of subject groups making a total of 351 for the three group comparisons.

Table V shows the frequencies, ex-

TABLE IV. Items on the Bender Recall Differentiating between Organics, Schizophrenics, and Normals, at or beyond the .10 Level of Confidence

Design and Item	Percent Recall Occurrence			
	Organics	Schizophrenics	Chi Square*	P
A, 10. Double line (2 or more)	22	3	2.90	.10
A, 14. Square rotated	22	3	2.90	.10
4, 4. Curls	60	0	4.22	.05
5, 1. Asymetry	28	3	4.61	.05
5, 10. Workover	32	3	5.91	.02
6, 1. Asymetry	60	33	3.90	.05
7, 8. Guide lines	0	24	3.38	.10
7, 9. Second attempt	11	38	2.76	.10
8, 2. Angles extra	48	12	10.90	.001
8, 5. Double line (2 or more)	24	5	4.06	.05
	Organics	Normals		
A, 19. Design missing	22	0	4.70	.05
5, 10. Workover	32	8	3.38	.10
6, 3. Point crossing	20	3	2.76	.10
6, 4. Curve extra	23	0	6.06	.02
7, 5. Double line (1 or more)	56	24	3.49	.10
8, 5. Double line (2 or more)	24	0	5.83	.02
	Schizophrenics	Normals		
A, 12. Tremor	16	43	3.54	.10
3, 5. Number of dots	18	83	7.33	.01
7, 9. Second attempt	38	3	8.51	.01
7, 8. Guide lines	24	0	5.84	.02

* Yates' correction for continuity was applied in all chi square computations.

TABLE V. Frequency with which Each Item on the Recall was Tabulated for Organics (O), Schizophrenics (S), and Normals (N), Expressed in Percent of Occurrence in the Designs Recalled for Each Group

Design A	O	S	N
1. Asym. circle	17	3	3
2. Asym. square	22	23	28
3. Break circle	0	0	0
4. Ends not join	9	0	0
5. Sq. not center	17	13	3
6. Sq. not join.	0	0	0
7. Sq. overlap	0	10	7
8. Angles extra	17	10	21
9. Ext. dot scat.	9	0	3
10. Double lines			
a. 1 or more	35	32	24
b. 2 or more	22	3	10
11. Guide lines	0	3	0
12. Tremor	26	16	43
13. Workover	0	0	0
14. Sq. rotation	22	3	17
15. Sec. attempt	4	16	7
16. Rotation	13	10	3
17. Distortion	9	6	10
18. Relative size	0	0	0
19. Design missing	22	13	0
No. Recalling Design	23	31	29

Design 1	O	S	N
1. Wavy line	29	20	21
2. Dot dash ci.	43	31	50
3. Dashes	10	0	0
4. Circles	5	3	0
5. No. dots (1 or more)	14	10	7
6. Double row	5	0	0
7. Workover	14	15	14
8. Sec. attempt	10	5	0
9. Rotation	0	3	0
10. Des. missing	0	3	0
No. Recalling Design	21	39	28

Design 2	O	S	N
1. Wavy line	62	41	44
2. Dots or dash	15	6	8
3. Shape circle	77	68	56
4. Ci. miss ext.	15	6	4
5. Ci. touch	0	0	0
6. Dev. slant	8	3	16
7. No. columns (1 or more)	15	9	12
8. Fig. 2 lines	0	0	0
9. Guide lines	0	0	0
10. Workover	8	6	0
11. Sec. attempt	0	0	0
12. Rotation	0	3	0
13. Des. missing	8	3	0
No. Recalling Design	13	34	25

Design 3	O	S	N
1. Asymetry	60	27	67

TABLE V. (*continued*)

Design 3 (cont.)	O	S	N
2. Dot dash ci.	60	45	67
3. Dashes	0	0	0
4. Circles	0	0	0
5. No. dots	60	18	83
6. Extra row	0	0	8
7. Blunting	0	0	0
8. Distortion	10	0	0
9. Guide lines	0	0	0
10. Workover	30	18	8
11. Sec. attempt	0	0	0
12. Rotation	0	9	0
13. Des. missing	50	18	17
No. Recalling Design	10	11	12

Design 4	O	S	N
1. Asym. curve	60	10	29
2. Break curve	0	0	0
3. Crv. not cent.	40	20	43
4. Curls	60	0	0
5. Not joined	0	20	0
6. Crv. rotation	0	30	14
7. Touch-up	0	0	0
8. Tremor	40	20	43
9. Distortion	40	40	57
10. Guide lines	0	0	0
11. Sec. attempt	20	10	14
12. Rotation	0	10	43
13. Des. missing	20	20	29
No. Recalling Design	5	10	7

Design 5	O	S	N
1. Asymetry	28	3	8
2. Dot dash ci.	68	41	61
3. Dashes	0	7	0
4. Circles	4	3	4
5. Ext. join dot	36	55	46
6. Ext. rotation	12	3	4
7. No. dots	8	7	4
8. Distortion	8	7	4
9. Guide lines	0	0	0
10. Workover	32	3	8
11. Sec. attempt	4	0	0
12. Rotation	4	0	0
13. Des. missing	12	17	8
No. Recalling Design	25	29	26

Design 6	O	S	N
1. Asymetry	60	33	42
2. Angles	30	15	16
3. Pt. crossing	20	13	3
4. Crv. extra	23	5	0
5. Double line (1 or more)	20	23	16
6. Touch-up	0	0	0
7. Tremor	37	44	42
8. Distortion	7	10	0
9. Guide lines	0	3	0
10. Workover	0	0	0
11. Sec. attempt	3	10	0
12. Rotation	3	0	3
13. Des. missing	10	5	6
No. Recalling Design	30	39	31

Design 7	O	S	N
1. Ends not join	0	10	3
2. Angles extra	67	38	38
3. Angles miss.	6	7	7
4. Ext. scatter	11	3	0
5. Double line			
a. 1 or more	56	41	24
b. 2 or more	22	24	10
6. Tremor	28	41	24
7. Distortion	44	41	21
8. Guide lines	0	24	0
9. Sec. attempt	11	38	3
10. Rotation	78	59	55
11. Des. missing	39	14	24
No. Recalling Design	18	29	29

Design 8	O	S	N
1. Ends not join	0	0	0
2. Angles extra	48	12	25
3. Angles miss.	0	0	0
4. Ext. scatter	6	2	0
5. Double line			
a. 1 or more	36	30	21
b. 2 or more	24	5	0
6. Tremor	36	37	32
7. Distortion	24	12	7
8. Guide lines	3	0	0
9. Workover	3	0	0
10. Sec. attempt	18	14	7
11. Rotation	0	0	0
12. Des. missing	6	5	0
No. Recalling Design	33	43	28

TABLE VI. Correlation Between Sums of Scores of Recalled Designs and Sums of Scores of Same Designs Copied

Group	N	Correlation Coefficient	P
Organics	33	.696	.001
Schizophrenics	50	.504	.001
Normals	33	.597	.001

TABLE VII. A Comparison Between Sums of Scores of Recalled Designs and Sums of Scores of Same Designs Copied

Group	N	t	P
Organics	33	5.66	.001
Schizophrenics	50	5.96	.001
Normals	33	7.11	.001

pressed in percent, with which each of the indices occurred in the recalls of the three groups. This tabulation is presented primarily as normative data for the present recall quantification obtained with three widely differing groups of subjects. In addition it provides an indication of the sorts of distortions to be expected on the recall regardless of the group studied. It may be noted that out of the total of 117 indices, 32 were not scored for the organics, 33 never occurred in the schizophrenic group and 42 were not found in the normal recalls. Of this number, 22 were absent in all three groups. The comparative superiority of the normals in this respect is emphasized by the fact that they recalled considerably more designs than either of the other two groups and yet made fewer kinds of errors.

Relationship between Recall and Copy Quality

A series of computations were performed to determine the relationships between the overall quality of the recalled designs and the accuracy of the original reproductions. This analysis revealed very significant correlations between the sum of the recalled design scores and the sum of the scores for the same designs copied. As shown in Table VI, this finding applied to all three groups with the highest correlation obtained for the organic group and the lowest for the schizophrenics. While the correlations are significant, it is to be observed that they are not sufficiently high to be used for individual prediction.

When the patients' and normals' recall and copy performances on the same designs were compared employing t to test for significant differences, it was found that the quality of the recall in all three groups was markedly inferior to the copied reproductions. The t values of the recall-copy comparisons are reported in Table VII.

DISCUSSION

The major purpose in the proposed investigation has been to subject the recall productions of the Bender-Gestalt Test to the same sort of rigorous quantification procedures that have thus far been applied only to the standard copy phase. The results of this study indicate that the P&S weighted score items for Designs 1 through 8 may be used effectively to quantify the recalled designs. In addition, the series of 19 indices developed for scoring Design A were found to be as reliable as the original P&S indicators and easily employed with both the copy and recall procedures. It is the authors' feeling that this modified P&S quantification is as comprehensive when applied to the recall as the original P&S system is when used with the copy phase. While 22 items out of a total of 117 recall indicators were not scored for any of the subjects included in this study, it has been the authors' experience that a comparable number of items are as rarely encountered in the copy productions. The number of subjects included in this study is insufficiently large to permit an adequate evaluation of whether certain items are entirely appropriate for the recall. Cross validation studies dealing with a variety of subject populations should clarify this point.

With regard to the application of the identical weights to deviations occurring in the recall as occur in the copy phase, again the number of subjects included in this study, as well as the nature of the control group, do not permit any definite statement on this issue. However, generally speaking, the present weights appear to be valid, in the sense of indicating the degree of stimulus deviation shown, when they are applied to the recall. For the most part the results of the group comparisons of summed weighted scores for each design are in agreement with what was anticipated; namely, that the organics tend to re-

call the Bender-Gestalt designs with less accuracy than either the schizophrenics or the normals, with the normals being the most accurate in their reproductions.

When group comparisons are made on specific scoring items, a somewhat surprising finding was the rather sparse number of signs obtained which differentiated the normals from the schizophrenics. In some contrast to the results of Arazi (2), who found that schizophrenics had significantly greater number of major distortions, including contaminations, in their recalls than did the normals, the present study found only two items in which the schizophrenics were significantly worse than the normals, and these were relatively minor distortions. The fact that the normal group used in this study did not differ from the other two groups more markedly with respect to the presence of severe distortions in their recalled designs, suggests that some of our conceptions of what constitutes a deviant recall may have to be revised. Furthermore, the obtained correlations between the extent of disturbance evident on the copied productions with that seen on the recall suggests that a significant portion of the recall distortion is a function of the original drawings. A more detailed analysis of the relationship between the copy and the recall may clarify this relationship and thereby improve the diagnostic usefulness of the recall procedure. This analysis is contemplated in the future.

Summary

The present study was designed as a pilot investigation directed toward the development of a system for quantifying the Bender-Gestalt recall. After examining a number of scoring systems dealing with both the recall and the standard copy procedures, the system devised by Pascal and Suttell for scoring the copy phase was selected as the basic model for recall quantification. Since this method did not include a scoring system for Design A, a number of items were developed which appeared to encompass the majority of deviations usually encountered in Design A as well as corresponding closely with similar items already used in scoring the other eight designs. The resulting recall scale included the 98 items used by Pascal and Suttell in their original quantification of Designs 1 through 8 together with an additional 19 items for Design A.

The Bender-Gestalt copy and recall procedures were administered to two patient groups, one having a variety of organic diagnoses and the other having solely schizophrenic diagnoses, and one normal group, all student nurses. The copy and recall protocols for each subject were then scored using the Pascal and Suttell method for the copy phase and the present modification of it for the recall. When the scores for the recalled designs were compared with similar scores for the same designs which had been copied, all three groups were found to be far worse in their ability to recall the designs accurately than in their ability to copy them acceptably. A design by design analysis using the P&S weighted scores revealed that, in general, the organics reproduced the designs least accurately from memory with the schizophrenics and normals coming next in this order. An analysis of the frequencies with which the specific scoring items occurred produced twenty indices which differentiated between the groups at the .10 level of confidence or better.

The results of this investigation suggest that the present quantification can be reliably applied to both the recall and copy productions, and that the 117 items comprising the total revised scale include the majority of deviations found in the recall. In that the number of subjects used in this research was not large, and there were certain limitations in the normal control group, cross valida-

tion studies are necessary to evaluate the appropriateness of the various recall scoring indices and their usefulness in discriminating between groups.

REFERENCES

1. Addington, M. C. A note on the Pascal and Suttell scoring system of the Bender-Gestalt Test. *J. clin. Psychol.*, 1952, *8*, 312-313.

2. Arazi, Shifra. Distortion in perceptual recall in schizophrenia. Unpublished doctor's dissertation, New York University, 1956.

3. Billingslea, F. Y. The Bender-Gestalt: an objective scoring method and validating data. *Clin. Psychol. Monogr.*, 1948, No. 1.

4. Bossom, J. The effect of certain variables upon recall: A study of the Bender-Gestalt Test. Unpublished master's dissertation, New York University, 1953.

5. Bowland, J. A., and Deabler, H. L. A Bender-Gestalt diagnostic validity study. *J. clin. Psychol.*, 1955, *12*, 82-84.

6. Gobetz, W. A quantification, standardization and validation of the Bender-Gestalt Test on normal and neurotic adults. *Psychol. Monogr.*, 1953, *67*, No. 6 (Whole No. 356).

7. Johnson, J. W. Diagnosis of the decompensated personality—its application to insurance medicine. *Proc. 39th Annu. mtg. med. sec., Amer. Life Convention*, 1951, 214-243.

8. Parker, J. W. Tactual-kinesthetic perception as a technique for diagnosing brain damage. *J. consult. Psychol.*, 1954, *18*, 415-420.

9. Pascal, G. R. Quantification of the Bender-Gestalt: A preliminary report. *Amer. J. Orthopsychiat.*, 1950, *20*, 418-423.

10. Pascal, G. R., and Suttell, Barbara J. *The Bender-Gestalt Test*. New York: Grune & Stratton, 1951.

11. Peek, R. M., and Quast, W. *A scoring system for the Bender-Gestalt Test*. Box 292, Hastings, Minnesota. Authors, 1951.

12. Reznikoff, M., and Olin, T. D. Recall of the Bender-Gestalt designs by organic and schizophrenic patients: A comparative study. *J. clin. Psychol.*, 1957, *13*, 183-186.

13. Robinson, N. M. Bender-Gestalt performance of schizophrenics and paretics. *J. clin. Psychol.*, 1953, *9*, 291-293.

14. Shapiro, M. B., Post, F., Lofving, B., and Inglis, J. "Memory function" in psychiatric patients over sixty, some methodological and diagnostic implications. *J. ment. Sci.*, 1956, *102*, 233-246.

15. Sonder, S. L. Perceptual tests and acute and chronic status as predictors of improvement in psychiatric patients. *J. consult. Psychol.*, 1955, *19*, 387-392.

16. Suttell, Barbara J., and Pascal, G. R. "Regression" in schizophrenia as determined by performance on the Bender-Gestalt Test. *J. abnorm. soc. Psychol.*, 1952, *47*, 653-657.

17. Tolor, A. A comparison of the Bender-Gestalt Test and the Digit-Span Test as measures of recall. *J. consult. Psychol.*, 1956, *20*, 305-309

Received March 12, 1957

An Experimental Variation of the Draw-A-Person Technique

EZIO PONZO [1]
Instituto di Psicologia, Universita di Roma

The draw-a-person technique devised by Machover (4) has already established itself as a useful projective device for the study of personality, and has been used in a number of valuable investigations. It is generally conceded, none the less, that the subject may be somewhat inhibited in his approach to the task, and that this fact may complicate problems of interpretation. The desirability of an experimental technique of "disinhibition," of producing a response set which might have the effect of "unmasking" some concealed elements, is therefore obvious.

A device which seems to serve this purpose rather well is that of asking S to draw a person "as an idiot would." While this instruction is in itself ambiguous (and therefore an invitation to projection), it constitutes also a suggestion to S that he draw in an uninhibited manner. The results reported here seem to support the view that this is exactly what happens.

PROCEDURE

The first step in the experimental procedure is to obtain two drawings of the human figure according to the standard instructions laid down by Machover. As soon as S has finished these drawings, the examiner presents the following comment and request: "The drawing of a human figure depends, among other things, on the intelligence of the person making the drawing. That is: an idiot, a stupid person, does not draw like a normal person. Now I would like you to do me the favor of drawing a person as you think you would draw it if you were an idiot. What I ask of you seems very queer, does it not? . . ." (A pause follows, long enough to give time to the examinee to agree with this comment by expressing his astonishment.) Then the examiner continues: "I can, however, give you some good advice: don't think about it too much, lay it out without worrying, as a real idiot would. The drawing is easier now than before; since you are drawing in the manner of an idiot, no one is going to criticize your product."

After the first drawing in Trial B (as an idiot; the normal drawing will be called Trial A), S is asked for another, of the sex opposite to that just sketched. In case the sex of the first production is unrecognizable, and is conceded to be asexual by S, two other figures are requested, following Machover's procedure.

Many S's protested at these instructions; they insisted on their absolute lack of knowledge as to the way in which idiots draw, and protested against the "difficulty" of the task. These reactions appeared to be mostly defensive in character. The examiner gave evasive responses to these protests, always returning to the original instructions. This procedure overcame the resistance in all cases. The Trial B drawings were actually produced more rapidly than those of Trial A.

Subjects. The present report deals with the drawings of 81 S's, 50 women and 31 men, inhabitants of the Latium region near Rome, age range 20-40 years.[2] All were well-educated, having at least the diploma from the *Scuola Media Superiore* (American high school level). None of the S's showed any overt psychopathology.

[1] The assistance of Ross Stagner, University of Illinois, with the English language version of this paper is appreciated.

[2] With the exception of one woman, aged 51.

TABLE I—Aspects of the Human Figure which are Generally Modified in Trial B as Compared with Trial A

	Increase	Diminution	Other Modifications
Size of the head	75.3%	4.9%	19.7%
Deviation of the upper limbs from the median line	58.7%	7.4%	34.9%
Relief of the extremities (feet and hands)	58.0%	8.6%	33.3%
Segmentation of the trunk of the feminine figure	14.8%	45.6%	39.5%

RESULTS

The drawings of Trial B have been compared with those of the same S on Trial A. Interest centers particularly upon *global aspects* of the drawing, *size* of the total drawing and of selected parts, *expressive style,* and *transformation of details.* The present discussion will be limited to those changes which are manifest by a substantial percentage of subjects.

Global aspects. The commonest trend noted in the drawings is toward simplification in Trial B. Segmentation of the trunk in the female figure decreases in 45 per cent of S's (Table I). Drawings are executed more rapidly, and many details are eliminated. Especially striking is the fact that 12 S's (14 per cent of the total) drew first on Trial B a figure without sexual differentiation. (This of course does not occur under the normal instructions used in Trial A.) Secondly, of the 35 S's who had drawn figures in profile or dorsal view—an "evasive" position—29 changed to a frontal view in Trial B. An example of this change is given in Figs. 1A and 1B, drawn by the same S.

The elimination of some details is accompanied by an exaggeration of others (data on size changes are given below). Strong relief is characteristic of the Trial B drawings; and there is little subordination of one part to another. The "surviving" details are not given equal weight in the drawing. For example, hands and feet are emphasized or exaggerated in Trial B (see Table I, and Figs. 6A, 6B).

Size. There is a tendency for the over-all size of the drawing to increase, but this is not uniform. In Table I it will be noted that there is usually an increase in the size of the head.

Expressive style. The importance of expressive style in relation to content has been emphasized by Bellak (1). In some cases the modification of style in Trial B is so extreme that it seems difficult to ascribe the drawings to the same author, as for example, Figures 2 and 3, all done by the same S. The Trial A drawings indicate uncertainty of line, absence of eyes, restrained posture, etc. The corresponding Trial B drawings add breasts to the female and genitals to the male

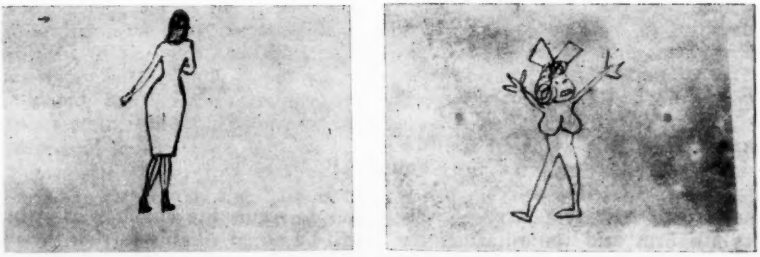

FIGURE 1. Drawings of S.G. (male), Trial A (left), Trial B (right).

FIGURE 2A

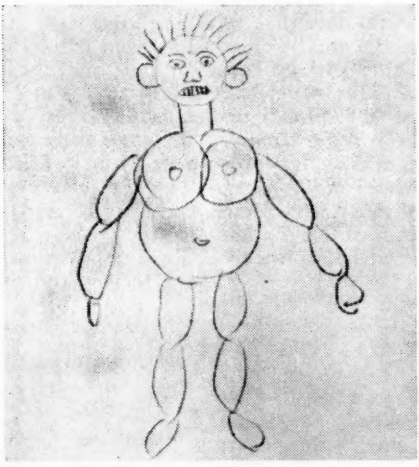

FIGURE 2B

figure, teeth and eyes to both; the lines are firm and unhesitating, the posture open and indeed aggressive. While it is difficult to quantify in simple terms the changes in style, it will be noted (Table I) that the position of the arms away from the body is highly characteristic of the Trial B

TABLE II—The Most Common Types of Hair-Drawing Associated with Male and Female Figures Drawn on Trial B

	Female Figure	Male Figure
Linear – descending hair	28.3%	1.2%
Linear – ascending hair	9.8%	23.4%
Absence of hair	9.8%	35.8%
Headgears	-- --	29.6%
Head with childish characteristics or attributes (ribbons, headbands, etc.)	9.8%	1.2%
Wavy hair, disheveled hair and other unusual drawings of hair	41.9%	8.6%

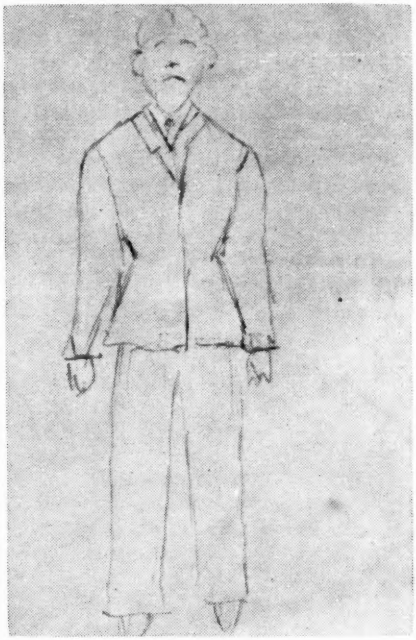

FIGURE 3A

drawings. The upward orientation of the arms (Fig. 1) is quite common. The addition of teeth, so striking in Figs. 2B and 3B, is rather frequent; 2 women out of 50, and 7 men out of 31, added teeth in Trial B. Even without teeth, the mouth may be emphasized. (Fig. 5.)

Treatment of hair. As various authors have suggested, the hair is an important detail in the drawing. Our S's show a marked tendency to accentuate the differences between male and female figures as regards the hair (Table II). Female hair is changed (where change can be depicted) to a disheveled appearance, or to descending pencil strokes, whereas male hair is represented more often by ascending strokes (Fig. 7B). Men, however, are quite likely to be represented without hair, or with hair concealed by some kind of headgear. Hair is shown as disheveled (Fig. 4B), with childish ornamentation (Fig. 1B) and so on.

Clothes. Changes with respect to clothing are of varied character. In

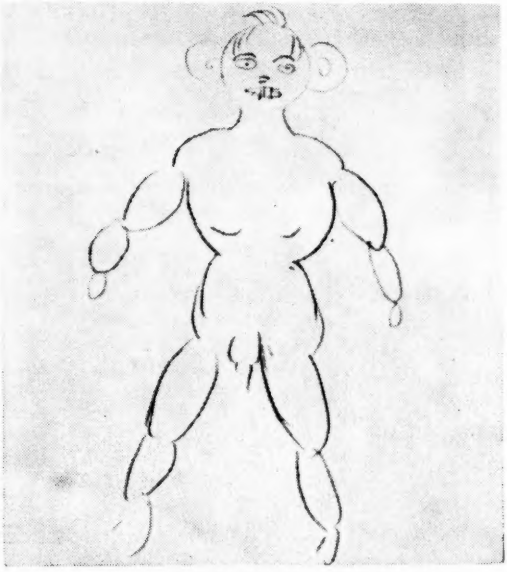

FIGURE 3B

FIGURE 4A (left), 4B (right)

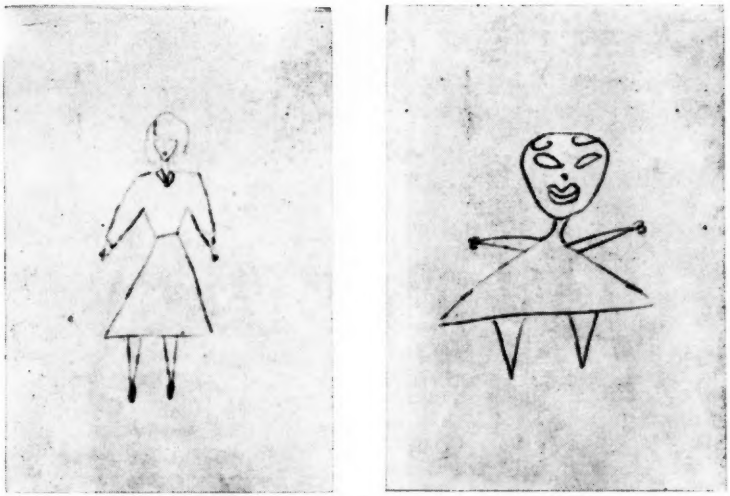

FIGURE 5A (left), 5B (right)

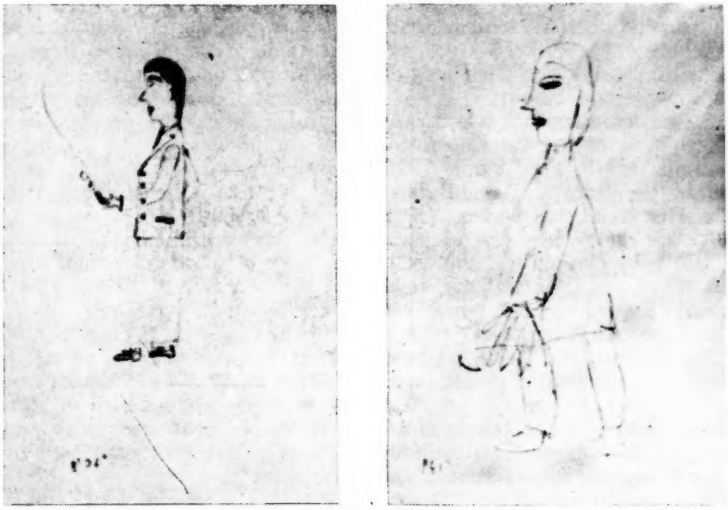

FIGURE 6A (left), 6B (right)

Trial B drawings the clothing may disappear; or a suggestion of conflict is presented when S draws a figure without clothes first, then traces clothing in afterward without bothering to erase the lines of the body which should be concealed by the clothing (Fig. 4B). This "transparency" seems related to other phenomena of a sexual nature. Some of these trends are as follows: in female S's, the breasts may disappear on Trial B, but if they

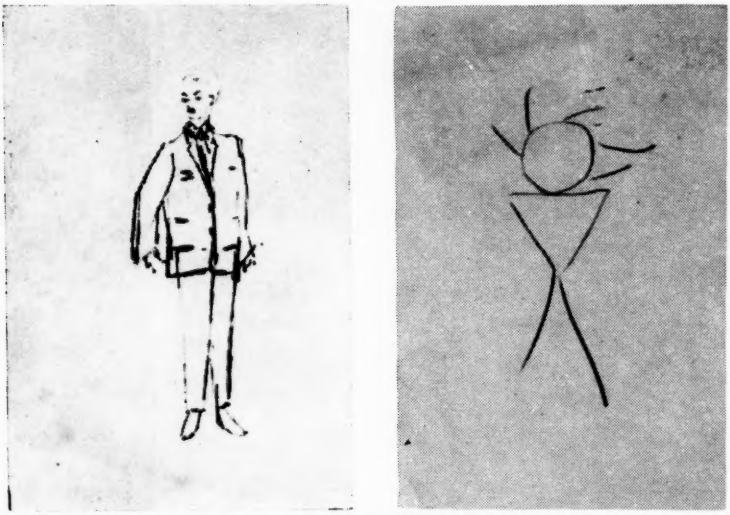

FIGURE 7A (left), 7B (right)

persist, they are accentuated; in male S's, the breasts are likely to be exaggerated on Trial B (41 per cent of males showed this trend).

The kinds of changes that occur, in a typical if somewhat extreme form, are summarized in Figs. 2 and 3, all drawn by the same S, a middleaged female. Her Trial A drawings (Figs. 2A, 3A) are characterized by extreme uncertainty, by marked segmentation, by absence of eyes (although eyebrows are clearly shown), by restrained posture and somewhat depressed expression. In contrast, the Trial B drawings are clear, muscular nudes, with added eyes, teeth, breasts and genitals; the lines are firm, the sketching rapid and careless. Both posture and animal-like teeth suggest aggressiveness. Indeed, the marked release of both sexual and aggressive symbolism in the Trial B figures reminds us of the often-postulated unity of aggressive and sexual impulses.

DISCUSSION

If we take these results at face value it is easy to conclude that the experimental variation in technique of asking a person to draw a human figure "as an idiot would" does have a disinhibiting or "unmasking" effect. However, it would be premature to consider this point proven. Because we are accustomed to assume that normal individuals have to some extent inhibited their sexual and aggressive strivings, we are not surprised when this experimental method elicits evidence of more outspoken urges in these directions. A more methodical check, however, would require clinical studies of individual cases in relation to the type of change manifested from Trial A to Trial B. What differences, for example, characterize subject SM (Fig. 5) as contrasted with subject MG (Fig. 2)? Such data are not yet available.

Another question which must be asked relates to the subjective attitude or "response set" of S while making these drawings. The use of the term *deficiente* ("idiot") necessarily gives rise to some ambiguity. The term does, nevertheless, carry some normal meaning. Therefore the manner in which S deals with this situation may be considered, as Cattell (2) has sug-

gested, a variety of dynamic distortion. This may be a case in which S perceives the idiot as a hypersexual and hyperaggressive person, and reflects this perception in his drawing.

It may be claimed, on the other hand, that S has only revealed his *stereotype* of an idiot; that is, if he believes that idiots are all sex and aggression, perhaps he projects these characteristics into the drawing. This would not argue against the validity of the draw-a-person technique, but it would cast doubt on the hypothesis that our experimental device produces disinhibition.

In favor of the "disinhibition" hypothesis, we can point to various aspects of the evidence which do not depend on the introduction of manifest symbolism. We note, for example, the signs of greater assurance in graphic style, increases in total size and in head size, reduction of total time required, elimination of evasive (profile and dorsal) positions, removal of the arms from the mean line of the trunk, greater relief of limbs and extremities, reduced segmentation of the trunk, etc. It is somewhat difficult to assume that all of these derive from a stereotyped conception of an idiot or an idiot's drawing.

One aspect of the "disinhibition" which seems fairly clear from comments by the subjects is that of freedom from self-criticism in the mechanics of drawing. It is as if S said to himself, "Now it is not important if I draw badly or not." This attitude was, of course, encouraged by the phrasing of the experimental instructions.

SUMMARY

An experimental variation of the Machover draw-a-person technique is reported. After producing drawings in the normal manner, S was asked to draw human figures "as an idiot would." Changes in the direction of simplification, exaggeration of sexual and aggressive details, and expressive style suggested that the technique had elicited a disinhibition, revealing aspects of the personality which were kept in concealment when drawing under the usual instructions. It is suggested that the technique represents a useful extension of the draw-a-person method as usually employed.

REFERENCES

1. Bellak, L. On the problems of the concept of projection. (In: Abt, L. E., and Bellak, L. *Projective psychology*. New York: Knopf, 1950.)

2. Cattell, R. B. Principles of design in "projective" or misperceptive tests of personality. (In: Anderson, H. H., and Anderson, G. L. *An introduction to projective techniques*. New York: Prentice-Hall, 1951.)

3. Gemelli, A. Percezione e personalita. *Arch. Psicol. Neurol. Psichiat.*, 1951, *12*, 475.

4. Machover, K. *Personality projection in the drawing of the human figure*. Springfield, Illinois: Charles C. Thomas, 1949.

Received March 15, 1957

The Validity, Bases, and Process of Clinical Judgment, Using a Limited Amount of Projective Test Data

WILLIAM T. POWERS AND ROY M. HAMLIN
Western Psychiatric Institute, University of Pittsburgh

In approaches to validation of the Rorschach test, it has been difficult to formulate research designs in such a way that the research mirrors the practical, everyday operations of the clinician. In an effort to handle this problem, researchers in one approach have used the judgments of the clinician as their data, focusing primarily on the end result of the integrative process rather than on the process itself. This raises a second problem, namely, how will the judgment of the clinician vary as a function of training, skill, the data available to him, etc. A third problem has to do with the existence and adequacy of criteria against which predictions from projective test data shall be compared, while a further problem concerns the type of behavior that we are trying to predict. Much of the research on the Rorschach test has been concerned with the exploration of relationships between Rorschach performance and overt behavioral reactions. Korner (1), however, states that the overt behavior of the individual is multiply and complexly determined, and that we are unable to effectively predict reality behavior from projective test data. She points out that this problem is not unique for clinical psychology, that it has not yet been solved by clinical psychiatry, and that it may never be adequately solved.

If prediction of reality behavior from test data has met with only minimal success, the question arises as to just what we are going to correlate with or predict from projective test data. The constructs developed in personality theory may have relevance at this point, intervening as they do between the needs of the individual and his overt behavior, even though perfect knowledge of these intervening variables in a given individual will not result in perfect prediction of reality behavior because of the fact that other factors will also be of importance in the determination of the emergence of specific reality behavior.

Individual case studies, as reported in the literature, have unsystematically followed this general approach. A recent article by Symonds (2) limited the amount of data presented to the judges to a single Rorschach protocol, sought to determine what aspects of the Rorschach material clinicians responded to in making judgments, and sought some measure of the validity of the clinicians' judgments by comparing them primarily with data obtained in a limited number of therapy interviews. Concern with the aspects of Rorschach material responded to by the judges represents an advance, but the unstructured nature of the judgment task in this instance makes it difficult to compare one judge with another and gives too much leeway to the judges in choosing areas of discussion.

The present study was designed to study the judgment process of clinicians in dealing with a small unit of projective test data. The authors decided to limit the amount of data presented to the judges for two reasons: (1) in order to make the task a less cumbersome one for the judges; and (2) in order that the judges might more easily specify the bases of their judgments in terms of specific test data. The unit of projective test data consisted of the responses to Card I of the Rorschach by a 34 year old male subject who was being seen in therapy as an out-patient at the

time he was tested. Information was available about the subject from 15 therapy interviews, a battery of test data, social service interviews with his family, and some follow-up data over a six-year period.

The subject's responses to Card I are as follows:

CARD I

19"

1. Something like a pelvis—view from bottom toward top.
2. Two animals with backs to each other—wolf nature.
3. A mask for a face. Ends for over the ears. Large slits: mouth, and chin.

96"

1. (whole response) General contour, from bottom up. The holes should not be there.
2. (whole response. He manages to fit in the whole blot, F—) Grotesque. Snout out of proportion (side wings are the "snouts").
3. (whole response)

METHOD

All of the judges used in the present study had had at least five years of experience in Rorschach administration and interpretation and all had a Ph.D. degree. Four judges were used in the initial phase of the study, all of whom had received the Ph.D. degree from the University of Pittsburgh. Six judges were used in the second and final phase of the study, all of whom had received a degree from different universities.

In the initial phase of the study the entire Rorschach of the subject was presented card by card to the four previously described judges and they were asked to record comments about the subject. This task was completely unstructured. The purpose of this procedure was to obtain a number of variables which would serve as common points of discussion for the six judges whose task it was to evaluate the performance of the subject on Card I, and to obtain variables which were not arbitrarily determined.

On the basis of the discussion of the entire Rorschach by these four judges, the following 11 variables were extracted: (1) intellectual level; (2) intellectual efficiency; (3) conflict areas; (4) defenses; (5) emotional control; (6) way in which subject relates to others and attitudes toward others; (7) self concept and self attitudes; (8) type of symptomatology; (9) diagnostic category; (10) identification; and (11) anxiety level.

In the second and final phase of the study, a mimeographed copy of the subject's responses to Card I was presented to each of the six judges, with the additional information that the subject was a 34 year old male. The judges were also given a list of the 11 variables listed above and they were asked to check those variables which they were willing to discuss on the basis of the data available to them. They were instructed to discuss each variable which they checked as completely as they felt necessary, but they were asked to put down their evidence for each statement in terms of specific aspects of the subject's performance on Card I.

RESULTS

The results of the present study will be considered from three aspects, namely: (1) validity; (2) the process of judgment; and (3) the general classes of cues used by the judges.

Validity. Insofar as validity is concerned, the impressions about the subject offered by the judges on the basis of the performance of the subject on Card I were compared with the impressions about the subject offered by a psychiatrist, who was unfamiliar with the results of the present study. His impressions were based on all of the available material and he discussed the same variables, with the exception of (1) intellectual level, and (2) intellectual efficiency. He felt

that the test data was more appropriate to these two variables.

All six judges estimated the intellectual level of the subject and the consensus was that the subject had high average intelligence, in the 110-120 I.Q. range. The subject's verbal I.Q. on the Wechsler-Bellevue, Form I was 124, his performance I.Q. was 103, and his full scale I.Q. was 115.

All six judges felt that the intellectual efficiency of the subject was impaired, that he had functioned at a higher level in the past. In evaluating the total test performance of the subject, with the life history and therapy data available to him, the psychologist who had tested the subject stated that his "intellectual functioning was uneven. Temporary inefficiency due to alcoholism undoubtedly played a part ... but some degree of gradual permanent loss must also be assumed."

In commenting about conflict areas, five judges stressed a sexual conflict, with homosexuality, latent homosexuality, and fear of heterosexuality being stressed. Three judges also stressed conflict over hostility, stressing either oral aggression, fear of his own hostile impulses, or conflict over the expression of hostility. The psychiatrist emphasized the following points of conflict: (1) passive-receptive needs which are frustrated, with resultant rage which must be repressed because of dependence on others; (2) conflict around aggressive, hostile wishes; and (3) intense guilt, with the patient making atonement by self-destructive tendencies.

In commenting about the subject's defenses, three of the judges mentioned three defenses while the remaining three judges mentioned two defenses. The defenses of suppression and repression were mentioned as a duo by three judges, suppression and somatization by three judges, and suppression and projection by two judges. These were the only defenses mentioned by more than one judge. Two defenses were mentioned only once, both by the same judge, and these two defenses were intellectualization and withdrawal. The psychiatrist mentioned five defenses, namely, regression, repression, denial, projection, and intellectualization.

In discussing the emotional control of the subject, the six judges variously stressed the brittle emotional control of the subject, his difficulty in maintaining control over hostility, and his overcontrolled emotionality. The psychiatrist stressed the fact that the subject defends strongly against affect, repressing anger and aggression and fearing sexuality.

In discussing the way in which the subject relates to others and attitudes toward others, all six judges stressed the fact that he is probably guarded and suspicious in his contacts with others, that he is threatened by them. The psychiatrist stressed the fact that the subject lets others assume responsibility and says in effect "take care of me." He relates in a passive dependent way to others, is chronically frustrated and therefore always enraged and angry, with the anger being repressed.

In discussing self concept and self attitudes, four judges primarily stressed the subject's feelings of inadequacy, incompleteness, and insufficiency. The psychiatrist stressed the fact that the subject feels cheated and deprived, and that he has chronically low self esteem and feelings of weakness and emptiness along with continuous superficial attempts to deny this.

In discussing type of symptomatology, one judge stressed probable somatization, poor work habits, and alcoholism or irregular eating habits, while a second judge stressed vague bodily symptoms, generalized anxiety, and withdrawal. The psychiatrist stressed the following symptoms: (1) self destructive tendencies; (2) passive dependency with oral demandingness; (3) depression; (4) rage which is largely repressed but shows itself in the sadistic domination of his environment; and (5) alcoholism.

In commenting on diagnostic category, one judge very tentatively diagnosed the subject as a schizophrenic with strong resistive and paranoid elements, while a second judge made a diagnosis of anxiety hysteria and felt that the subject was moving toward a schizophrenic break. The psychiatrist diagnosed the subject as a personality pattern disorder, a passive dependent personality with chronic alcoholism.

Three judges discussed the subject's identification and they stressed variously his confused identification, his questionable identification with a male figure, and the fact that underneath he is desirous of adopting a passive role even though his conscious identification is generally probably masculine. The psychiatrist felt that the subject's identification was primarily with his mother in a passive dependent way.

All six judges agreed that the subject's anxiety level was high, with one judge commenting on the difference between "manifest" anxiety and "underlying" anxiety. The psychiatrist stated that the degree of repression and other defenses indicate very great anxiety.

Because of the limited amount of data available to the six judges and because of the lack of specific definition of the 11 variables, we would not expect perfect coincidence of judgments with the criterion measures. A further deterrent to perfect agreement is the fact that the psychiatrist dealt with a much larger amount of material and, out of economy or necessity, might tend to highlight certain trends at the expense of others. The results bear this out. Agreement with the criterion measures was fairly good for the variables of intellectual level, intellectual efficiency, self concept and self attitudes, identification, and anxiety level; and agreement was least adequate for the variables of way in which subject relates to others and attitudes toward others, type of symptomatology, diagnostic category, and emotional control. In the case of the variable of type of symptomatology it seems obvious that the differences between the judges and the psychiatrist arise in large part from the particular factors which each feels should be subsumed under this term, and all of the symptoms pointed out by the two judges are actually consistent with the overall clinical picture. The disparity between the judges and the psychiatrist in handling the variable of emotional control appears in large part to be a function of a difference in emphasis, with the judges primarily emphasizing the ineffectiveness of the subject's control of impulses while the psychiatrist primarily emphasizes the subject's strong need to control these impulses.

The use of the 11 variables as common discussion points appears to have been a worthwhile modification, focusing attention not only on interjudge variation but also on the variation in the type of judgments made by the judges as a group and by the psychiatrist. As an example of the latter, the listing of alcoholism as a symptom is a simple matter of looking at the historical data for the psychiatrist, whereas it is a complex inference on the part of the psychologist and is offered tentatively.

The 11 variables were not spelled out or delimited in any way for the judges or the psychiatrist, and the inadequacy of this procedure is apparent, at least for many of the variables, because of the leeway given to the judges in their interpretation of these variables. The more different such interpretations are, the more difficult inter-judge comparisons become, as well as comparisons between their judgments and the criterion measures. In any future study of this type, the variables should be delimited.

The process of judgment. In the process of arriving at their judgments, the judges made inferences on the basis of either: (1) several units of evidence; (2) all of the available

evidence; or (3) one unit of evidence. They used the above procedures in descending order of frequency. Two other procedures were used in making inferences, but with practically negligible frequency. In these two procedures the judges either: (4) made an inference on the basis of response sequence; or (5) made an initial inference on the basis of one or more units of evidence and then built a structure of further inferences, logically bound together and related to the initial inference, with the basis of the logic being psychoanalytic theory.

The general classes of cues used by the judges. The general classes of cues responded to by the judges, in descending order of frequency, were as follows: (1) content; (2) elaborative comments; (3) the overall pattern of responses; (4) traditional Rorschach scores; and (5) vocabulary. One judge also responded to (6) the response sequence.

In order to tie the process of judgment and the particular cues used by the judges to the preceding material on validity, an example will be taken from each of the 11 areas discussed under validity. Each example was chosen because it is illustrative of the primary type of evidence offered by the judges as a group in discussing each variable, and each example will be described briefly in terms of process and cues.

1. Intellectual level — One judge said that the subject had bright-normal intelligence because of (a) his vocabulary usage and (b) his W tendency. The judge therefore used several units of evidence, namely, vocabulary and traditional Rorschach scores.
2. Intellectual efficiency—One judge said that at one time the subject "probably functioned somewhat higher than he does now because the response pattern is a 'sick' one and would imply that energy needed for intellectual pursuit or creativeness is 'tied up' in maintaining control." The judge used all of the available evidence, the overall pattern of responses, in arriving at a decision.
3. Conflict areas — One judge said that "a homosexual conflict is suggested because of the subject's manner of perceiving the pelvis, i.e., the view from the bottom toward the top." The judge used several units of evidence in this instance, namely, the pelvis content and the subject's elaboration of this content. After his initial inference, this judge continued, "At a deeper level this homosexual conflict might indicate a conflict over whether to relate passively or actively; in turn, this points to deep dependent needs and probable early frustration of security." The latter comments illustrate the process of building a structure of inferences, and is one of the two instances of the use of this procedure.
4. Defenses—One judge said that the subject used suppression as a defense because of the "constriction and guardedness implied by the 'mask' response and by the absence of other than form responses." In this case the judge made use of several units of evidence, namely, specific content and traditional Rorschach scores.
5. Emotional control — One judge said that the subject's "emotional control is poor because of the F— response and its content, i.e., the distortion of the blot into something predatory and destructive." He thus made use of several units of evidence, namely, traditional Rorschach scores and specific content.
6. Way in which subject relates to others and attitudes toward others. One judge said that the subject "is fearful of close contact because the back to back stance of the animals suggests inhibition and avoidance along with evasion." The judge primarily

made use of a single unit of evidence, namely, the specific elaborative comment concerning the animal response.

7. Self concept and self attitudes — One judge said that the subject "has strong feelings of inadequacy" because the 'mask' content suggests that he "covers up." In this instance the judge used one unit of evidence, the mask content, with his inference that the subject has feelings of inadequacy being based on his previous inference that the subject's need to "cover up" betrays the presence of underlying feelings of inadequacy.
8. Type of symtomatology — One judge mentioned "withdrawal from others" as one symptom. The judge said that the "grotesque" comment suggests that the subject "feels different from others, i.e., odd" and that he withdraws from others because of this. The judge therefore used one unit of evidence in this instance, an elaborative comment by the subject, inferring that this comment has a self reference and inferring further that those people with such a self concept will tend to withdraw from others.
9. Diagnostic category—One judge said that "although I feel that I should have considerably more evidence than this, I do get the impression of schizophrenia with strong resistive and paranoid elements. Another way of putting it—I have seen schizophrenics perform in much this way." In this case the judge was making use of all of the available evidence, the overall pattern of responses, in arriving at a decision.
10. Identification — One judge said that the subject's "conscious identification is generally masculine" because of the "wolf nature" response, but felt that underneath the subject is quite desirous of adopting a passive role because of the "pelvis, view from bottom toward top" response. The judge therefore used several units of evidence, namely, content and elaborative comments.
11. Anxiety level—One judge said that the subject's anxiety level was high "because the pelvis response reflects bodily concern, the wolf nature response reflects feelings of threat from without, and the mask response reflects a necessity to protect his feelings from the view of others." The judge used several units of evidence, all of which may be labeled content, and the inference about anxiety level was based on the accumulation of specific inferences about each element of content.

The inferential nature of clinical judgment is quite in evidence in the above examples, with the judges "taking off" from the small amount of data available to them and speculating far beyond the most obvious characteristics of these data. The inferences of the judges appear to be primarily a function of the following factors: (1) personal experience with the Rorschach test, with the subsequent development of sets and expectations, some of which are personalized but most of which are held in common; (2) familiarity with personality theory in general and the ability to use this knowledge in Rorschach interpretation; and (3) the ability of the judge to actively formulate hypotheses about what the subject is like, with this ability probably being a function of either the motivation of the judge or the general clinical acumen of the judge or, most likely, both factors.

All of the judges had roughly had a similar amount of clinical experience. Several judges, however, were more "successful" than the others in approximating the criterion measures, and their greater degree of success may be attributed primarily to their superior ability to make use of their knowledge of personality theory in

making inferences from the Rorschach.

The primary cues used by the judges were overwhelmingly what might broadly be referred to as "content," as opposed to traditional Rorschach scores. The judges rarely made statements on the basis of Rorschach scores alone, using them primarily as supportive evidence. Undoubtedly, however, the judges were implicitly taking such factors into consideration, since they were undoubtedly taking the subject's total card performance as their base.

Discussion

The encouraging degree of validity obtained in the present study suggests that it may be profitable to pursue the methodology used, with modifications. The eleven variables, for the most part, pertain to personality factors which intervene between the needs of the individual on the one hand and his overt behavior on the other hand, and are consistent with Korner's emphases. There is nothing sacred about the present list and they may well be modified, either by definition, deletion, or addition. In terms of obtaining maximal cooperation from the judges it might be well to limit the number of variables to less than ten.

The general attitudes of the judges in approaching the present task might be summed as follows: (1) they accepted the subject's handling of Card I as a valid sample of the handling of a task with which they had had extensive clinical experience, namely, Card I of the Rorschach; (2) they assumed that *what* the subject perceived, the *manner* in which he perceived it, and the way in which he *described* his responses were not chance matters but were instead a function of his personality pattern; and (3) they felt that one aspect or another of the limited sample of behavior made available to them was pertinent in making statements about the subject with regard to his intellectual level, conflict areas, etc.

In the framework of this general attitude it is possible to undertake a study of the judgment process of clinicians in two ways. The judges in the present study apparently made initial judgments about the more or less obvious characteristics of the behavior sample confronting them. These initial judgments are not always explicitly spelled out by the judges, but are obviously close to a descriptive level whenever this is done. Marked disagreements between the judges were rare at this level, though disagreements did occur, for example, concerning the quality of the whole responses of the subject, the degree of elaboration of the subject's responses, and adequacy of the number of responses given by the subject to Card I. It should be relatively easy to ferret out disagreements at this level of judgment. The judges then "take off" from these initial judgments to make further speculations. These second level judgments may be examined both from the aspect of logic and from the aspect of validity, in terms of some criterion such as the one used in the present study.

Summary

The present study investigated the validity, bases, and process of clinical judgment, using a limited amount of projective test data, namely, Card I of the Rorschach of a 34 year old male subject. The study introduced a modification of having the judges discuss the same factors with respect to the subject, instead of allowing the judges complete freedom in deciding what to discuss. The results show that judges are able to make reasonably valid statements about a subject on the basis of a limited amount of projective test data. The procedure also shows the necessity of fairly exact delimitation and specification of the variables such as defenses, etc., if meaningful inter-judge comparisons are to be made as well as meaningful comparisons of their judgments with a cri-

terion measure. The results further show that, within the framework of the subject's overall performance on Card I, the judges tended to single out factors which might broadly be referred to as "content" in support of their inferences rather than traditional Rorschach scores. Whenever the latter were singled out, they were generally used as supportive evidence rather than primary evidence. Finally, the study shows clearly the inferential nature of the judgment process, with the judges "taking off" and going far beyond the most overt characteristics of the data. The most important variables in this "taking off" process appeared to be personal experience with the Rorschach test, a knowledge of personality theory and the ability to apply this knowledge in Rorschach interpretation, and the general clinical acumen as well as the motivation of the judge. It was suggested that the judgment process consists of two levels, more or less, with the first level being at a more or less descriptive level and the second level using the initial judgments as a jumping off point for speculation. It was further suggested that disagreements can be fairly easily uncovered at the descriptive level, while the second level speculations may be examined from the aspect of logic and from the aspect of validity.

REFERENCES

1. Korner, Anneliese F. Limitations of projective techniques: Apparent and Real. *J. Proj. Tech.*, 1956, *20*, 42-47.

2. Symonds, Percival M. A contribution to our knowledge of the validity of the Rorschach. *J. Proj. Tech.*, 1955, *19*, 152-162.

Received May 24, 1957

The Prediction of Overt Aggressive Verbal Behavior from Rorschach Content[1]

GORDON E. RADER
Yale University[2]

The research reported in this paper represents an attempt to establish a relationship between the content of Rorschach responses and the overt behavioral expression of aggression. No assumptions are made concerning the degree to which the subject is aware that he is expressing aggression; a purely overt behavioral approach is intended. As the title suggests, only a limited type of aggressive behavior was under study here. The intent of the study was to determine whether or not the content of the subject's responses is related to his overt behavior in at least one type of situation. This study was limited to the content of responses as opposed to formal scoring categories for three reasons: because previous studies (8, 21) have failed to demonstrate any extensive relationship between aggressive behavior and the formal scoring; because some authorities (27) feel content to be the more fruitful approach in general; and, most important, because one of the aims of this study was to shed light on previous studies in this area which utilized content (6, 8, 9, 18, 21, 23, 25).

There have been a number of attempts to show a relationship between the content of the Rorschach and various measures of aggressive behavior or aggressive drive (6, 8, 9, 18, 21, 23, 25). These studies all assume a positive relationship between hostile content on the Rorschach and underlying aggressive drive, and the results generally confirm this. However, there is some conflict among the studies as to whether hostile content is positively (8, 9, 21, 23) or negatively (6, 18) related to *overt* aggressive behavior.

The present study differs from previous studies, both in the mode of aggressive expression to which prediction is made, and in the nature of the Rorschach content from which prediction is made. Thus, previous studies have either not attempted to predict specific overt behavior at all (6, 25), or have been primarily concerned with extreme physical expressions of aggression (8, 21, 23). Other studies (9, 18) did not involve explicitly defined aggressive behavior. The present study seeks to predict to a behavioral criterion, but one which is much less extreme in form than assaultive behavior and, also, which is on a verbal rather than a physical level. In addition, previous published studies have not taken into account the inhibitory or counter-aggressive forces in the personality, which may modify the extent to which aggressive impulses of any particular intensity will be expressed in behavior. From many theoretical positions (e.g., 5, 7, 17) these counter-aggressive forces are as important in determining the behavioral outcome as the original drive itself. The research reported

---

[1] This article is based upon a doctoral dissertation submitted to the Department of Psychology, Yale University, in partial fulfillment of the requirements for the Ph.D. degree. The author is indebted to the members of his doctoral committee, Dr. Seymour B. Sarason, Dr. Irving L. Janis, and Dr. George Mahl, for their guidance. The full cooperation of Dr. Lawrence Freedman and the Yale Study Unit in Psychiatry and Law and the administration of the Connecticut State Prison is much appreciated. Special thanks is due Dr. John Williams, Mr. Tom Davis Olin, and Mr. Donald Tomblen who assisted in the gathering and scoring of the data and offered helpful criticism and advice. Dr. Marvin Reznikoff and Dr. John Donnelly were kind enough to extend the facilities of the Institute of Living, Hartford, Connecticut, for the analysis of the data of this study.

[2] Now at University of North Carolina.

herein represents a more refined attempt to predict overt aggression from the Rorschach, by estimating Rorschach content indicators of both aggressive drive and inhibitory forces in the personality.

SUBJECTS AND PROCEDURES

The procedure consists essentially in correlating Rorschach content indices of aggressive potential with a criterion measure based upon the subjects' behavior in therapeutic discussion groups as rated by the therapist and an observer.[3]

Subjects:

The subjects for the study consisted of state prison inmates who were taking part in six weekly discussion groups at the prison. During the six months in which these groups met, 44 inmates remained in the groups long enough to provide usable data.[4] Some of the Rorschach protocols were extremely sparse, however, and it was felt that such records could not be used for valid prediction. A difference of only one in the incidence of a particular type of content in such a barren record would change all the prediction scores markedly. Therefore, an arbitrary ruling was made that no record whose self-expression score (see below) was 9 or less would be used in this study. This eliminated six subjects making the final number of subjects utilized in the study equal to 38.

An effort had been made to rule out mental defectives and psychotics and to choose "acceptable" candidates for psychotherapy. Although a disproportionate number of subjects in the groups were sexual offenders, the men in the groups represented a wide sampling of criminal behaviors (e.g., murder, housebreaking, armed robbery, carrying unlawful drugs, embezzlement, homosexuality, rape, statutory rape, and pedophilia). Therefore, the population used in this study was composed of relatively non-pathological cases, psychiatrically and intellectually, who have deviated markedly from the social code.

Four groups were conducted jointly by two psychiatric residents. These groups were more carefully screened with regard to the criteria of therapeutic prognosis than the remaining two groups which were conducted by a psychologist with the experimenter as observer. Because of the differences in criteria of selection, and because of differences in the criterion data, the two groups and the four groups were treated separately in the analysis. The psychologist's groups will be termed Sample A and the psychiatrists' groups will be termed Sample B.

Rorschach data

Rorschachs were all administered by the experimenter, before or during the first month of the group discussions. With the exception of four subjects who were observed once, all subjects were tested before being observed in the groups by the experimenter. The protocols were typed, coded, and filed until after the groups were disbanded more than six months later. They were then scored, without identifying information, by the experimenter. In addition, the records of Sample A were scored independently by one psychologist and the records of Sample B by a second psychologist, neither of whom had

[3] In the original design of the study, it was intended to get ratings or other evaluations from the prison personnel. However, this was not possible for administrative reasons. An attempt was made to use prison disciplinary reports as well, but this proved to be impossible because of the infrequency of such reports and because they tended to occur only during the first year or two until the inmate became "prison-wise" and learned to avoid trouble. No other prison records were found to be useful.

[4] Some of these men either left the prison on discharge, parole, or transfer or they were given jobs in the prison which conflicted with the group meetings. There is no reason to believe this was a selected sub-sample. The rest, about five men, left because they voluntarily decided they no longer wished to continue. The selective factor here is undetermined.

any connection with the prison or knowledge about the prisoners.

The Rorschach was scored in terms of content indicators of both aggressive drive and inhibitory forces in the personality. Underlying aggressive drive was assumed to be reflected in content depicting hostile behavior or feeling, interpersonal conflict, instruments or forces of violence and destruction, aggressive animals or persons, and evidence of accomplished aggression or destruction. Two content categories, termed Hostility and Mutilation (H and M), subsume the contents presumed to reflect underlying aggressive drive. The final expression of this drive is assumed to be dependent upon other personality forces. For example, a person who tends to see only positive or friendly attitudes about him, is likely to manifest these same attitudes to others, even though in some cases these tendencies may be dynamically related to aggressive drive. Similarly, one who seeks to win affection and approval by adopting a deferential, submissive attitude is more likely to curb the overt expression of aggression. Inhibition of aggression may result in a passive, unassertive, or inactive orientation, and in a desire to escape from the needs for assertive, aggressive behavior through regression to the passive-receptive state of childhood (7). Feelings of guilt derive from a strong superego, which is likely to be directed against the expression of aggression as well as other unacceptable impulses. Although the role of the mechanism of redirecting hostility against the ego in the etiology of depression is a matter of some dispute, it seems generally agreed that depressed individuals do tend to direct their hostility toward themselves rather than toward the environment (1, 5, 7, 26). Aggression is frequently inhibited in those who feel inadequate because they do not feel capable of standing up to others in an aggressive fashion and because they try to bolster their self-esteem by winning the approval of others (7). To attempt to provide some indication of the presence and strength of the foregoing tendencies, attitudes, and feelings, four further categories of content were constructed: Passivity or Friendliness, Guilt, Depression, and Inadequacy. Content was scored in these four categories when it included references to emotions, attitudes or behavior reflecting passivity, friendliness, guilt, depression or inadequacy. It also included references to animals whose cultural stereotype was docility, gentleness, frailty, weakness, or inferiority; reflections of sensitivity to aesthetic qualities; references to religion, purity, or punishment; references to death (without implying an aggressive agent), disease, ruin, decay, desolation, or coldness. The personality factors assumed to be reflected in these four content categories are not specifically directed against or indicative of the inhibition of aggression, but are factors which are considered to be frequently associated with the inhibition of aggression, though not always (e.g., guilt may arise from a sexual conflict, and passivity or inferiority feelings may be reacted against with aggressiveness). Hereafter, the content categories will be referred to by their initials (HM content being associated with aggressive drive and PGDI content being associated with the inhibition of its expression).[5]

Each Rorschach response was scored for another factor, called self-expression. Previous studies have used only the number of hostile responses, sometimes as a percentage of the total

[5] Many of the ideas for the content scoring schema employed in this study, as well as many of the specific examples used, were derived from Elizur (6), Schafer (20), and DeVos (4). The full schema for scoring content in these six categories has been deposited with the American Documentation Institute. To obtain this material, order Document No. 5193 from American Documentation Institute, 1719 N Street, N.W., Washington 6, D.C., remitting $1.75 for 35 mm. microfilm or $2.50 for 6 x 8 in. photocopies.

number of responses, but they failed to consider the subjectivity of the other responses. Bland, stereotyped, guarded, or popular responses are more objective, while unusual, idiosyncratic, emotionally invested responses are more subjective and the content therefore more expressive of some aspect of the self (e.g., motivation, attitudes, mood, defenses). The present study has sought to obtain a measure of the degree of self-expression or subjectivity in each response, based upon such factors as the normative frequency, form level, specificity, and emotional intensity of the response. The scale ranged from 0 to 3 as follows:

0=popular responses; or vague and evasive responses;

1=popular responses with unusual elaboration; or other responses which are not vague, original, of poor form level, or emotionalized;

2=mildly emotionalized responses which are neither original nor of poor form quality; or any unemotionalized responses which are original and/or have poor form quality.

3=strongly emotionalized responses; or any mildly emotionalized responses which are original and/or have poor form level.

It was felt that the total self-expression in the content of a record is a more useful base than the number of responses for calculating proportions of specific categories of revealing content, because it is a more accurate index of the total amount of interpretively meaningful content in the record. This permits one to distinguish between two records of equal length, one containing little hostile content and little content reflecting other drives or personality factors, the other record also containing little hostile content but having content which reveals much about other drives and personality factors.[6]

[6] A more complete description and instructions for scoring of this variable have been deposited with the American Documentation Institute. It may be obtained by ordering Document No. 5193 from American Documentation Institute, 1719 N Street, N.W., Washington 6, D.C., remitting $1.75 for 35 mm. microfilm or $2.50 for 6 x 8 in. photocopies.

Criterion Data

The criterion of aggressive behavior was based upon a rating scale.[7] At the end of the hour a global rating of the aggressiveness of the subject during that hour was made upon a 9-point scale. This scale ranged from very compliant, submissive, self-abasing behavior through a relatively neutral point to openly very hostile behavior at the other extreme. Most of the points of the scale were anchored by descriptions of the types of behavior characterizing that level, although the final rating could represent a subjective averaging of several levels of aggressive behavior. A high rating in aggressiveness was given to behavior which was primarily assertive, bitterly denunciatory of the prison administration, dominating, challenging, or which occasionally openly disregarded the feelings and status of others present. A low rating was given to the absence of this sort of behavior and the presence of a more generally compliant, submissive, self-depreciatory attitude, or one which was steadily friendly and positive.

In Sample A, the ratings were made on 11 sessions but because some of the subjects missed some of the sessions, an average of only 8.0 sessions was rated for each subject. In sample B, the ratings were filled out for 15 sessions, but, because of absences, this amounted to 10.5 sessions per subject.

[7] Since both these correlations involve one variable in common, the standard procedure for determining the significance of the difference between correlations could not be employed. A formula suggested for this purpose by Quenouille (15, p. 70) was used instead. The formula used was: $\frac{r_{xa} - r_{ya}}{2(1 - r_{xy})}$ and the result is to be treated as another correlation coefficient with n−2 degree of freedom.

In both samples, the number of sessions for which there was usable information ranged from two to the maximum for the different subjects. All subjects with less than two sessions represented on the criterion measure were discarded.

In Sample A, both therapist and observer (who was always the experimenter) rated each session. In Sample B, either the observer or the therapist, but never both, completed the ratings. During four sessions, however, either the experimenter or the psychologist who conducted the groups in Sample A replaced one of the psychiatrists as the observer in Sample B. During these four sessions, both therapist and observer made ratings. Thus a complete measure of inter-rater reliability is possible with the criterion measure in Sample A, but only a sample reliability check is possible in Sample B.

Because shortage of personnel necessitated that many of the observations be made by the experimenter, the possibility of contamination arises at two points. First, the experimenter's memory of the Rorschach protocols which he gathered may have influenced his observations. Secondly, his observations may have influenced his scoring of any protocols which he consciously or unconsciously recognized. The criterion data gathered by the experimenter were therefore excluded from the results and the Rorschach data were submitted to an independent scoring. Data involving the experimenter are used only for reliability determinations.

Rationale for Hypotheses

Two principal classifications of content were previously delineated: Hostility and Mutilation content (HM), which is assumed to be related to the presence of aggressive drive; and Passive or Friendly, Guilt, Depressive, and Inadequacy content (PGDI), which is assumed to be related to the inhibition of aggression. The aim of the present study was to determine if any relationship existed between these two types of content and overt verbal aggressive behavior. However, two possibilities exist with regard to such a relationship, which though not mutually exclusive, may lead to different predictions under certain conditions. HM and PGDI content may either be independently related to aggressive behavior, or these contents may be related to aggressive behavior primarily in terms of their relationship to each other. If they are independent, then the best estimate of the potentiality to act aggressively will be proportional to an appropriately weighted algebraic sum of the aggressive implications of HM content and the inhibitory implications of PGDI content. Multiple correlation performs this operation to optimal advantage. Moreover, accepting the opposing function of HM and PGDI with regard to the expression of aggression, an independent relationship implies that they must be oppositely correlated with aggressive behavior.

If, on the other hand, these contents interact in their relation to aggressive behavior, then the best prediction will be obtained by a function representing their interaction. One way in which these two types of content might be meaningfully interrelated is analogous to the balance of opposing tendencies in the conflict over the expression or inhibition of aggressive behavior. The ratio of content expressive of aggressive drive to all the content relating to either aggressive drive or inhibitory forces (i.e. HM/(HM + PGDI) might then be viewed as an index of the threshold to aggressive behavior. If this particular interaction were the crucial factor, then it would be possible for this function to be positively related to aggressive behavior irrespective of whether either, both, or neither of these two types of content were themselves positively related to the expression of aggression.

It might be remembered that self

expression (SE) was proposed as a more accurate index of the total amount of interpretively meaningful content in the record. It might be expected, then, that Rorschach content variables expressed in terms of SE, rather than R, would show a higher relationship to the criterion.

ALTERNATIVE HYPOTHESES

1. *HM and PGDI content are independently related to aggressive behavior.*

Predictions:

a. A positive correlation between HM /R and the criterion.
b. A negative correlation between PGDI/R and the criterion.
c. Maximal prediction is obtained with a multiple regression type equation.

2. *The relationship between HM and PGDI content is the crucial factor in relating this content to aggressive behavior and the ratio of HM content to HM + PGDI content is related to the likelihood that an aggressive impulse, when aroused, will be expressed rather than inhibited.*

Prediction:

A positive correlation between HM/(HM + PGDI) and the criterion.

3. *The proportions of HM and PGDI contents in the record are more accurately represented in relation to SE than in relation to R.*

Predictions:

a. HM/SE is more positively related to the criterion than HM/R.
b. PGDI/SE is more negatively related to the criterion than PGDI/R.
c. A higher relationship to the criterion is provided by a multiple regression type equation employing HM/SE and PGDI/SE than one employing HM/R and PGDI/R.

RESULTS

Predictions

The results pertaining to the specific predictions outlined in the Hypotheses section are presented in Table I.

Prediction 1a: A positive correlation between HM/R and the criterion (see Row 1). This prediction is supported by the results. The combined

TABLE I—Coefficients of Correlation between Rorschach Prediction Variables and the Aggressive Rating Scale Criterion with Combined Confidence Levels for Both Samples.

Rorschach Prediction Variable	Sample A (N=15)	Sample B (N=23)	Combined p-value*
HM/R	.34	.41	.01
PGDI/R	−.06	−.28	.15
HM/R and PGDI/R†	.41	.56	.01
HM/(HM+PGDI)	.25	.46	.02
HM/SE	.35	.47	.01
PGDI/SE	−.24	−.32	.05
HM/SE and PGDI/SE†	.45	.54	.01

Note: In Sample A, a correlation of .592 is needed for .01 level of confidence; .441 is needed for .05 level of confidence; .351 for .10 level of confidence; .235 for .20 level of confidence.
In sample B, a correlation of .482 is needed for .01 level of confidence; .352 is needed for .05 level of confidence; .277 for .10 level of confidence; .184 for .20 level of confidence.

* These values were obtained by combining the p-values of the results in each sample according to a method described in Mostellar and Bush (11).

† These coefficients were obtained by correlating the actual criterion scores with values predicted for the criterion from a multiple regression equation, derived in each case from the data of the opposite sample and then applied to the data of the same sample. The two Rorschach Prediction Variables indicated were the multiple predictors.

probability value reaches the .01 level of confidence.

Prediction 1b: A negative correlation between PGDI/R and the criterion (see Row 2). This prediction is not confirmed by results which reach an acceptable level of confidence, in either sample. The result in Sample B falls within the .10 level of confidence but, when the two samples are combined, the confidence level falls to .15.

Prediction 1c: Maximal prediction is obtained with a multiple regression type equation (row 3 greater than other rows except row 7). Sample B aggression ratings were predicted by means of the multiple regression equation computed on the Sample A data, and Sample A aggression ratings were predicted from the Sample B multiple regression equation. HM/R and PGDI/R were the two predicting variables. These predicted criterion scores were then correlated with the obtained criterion scores. The correlations so obtained are significantly higher than those obtained using HM/(HM + PGDI) (row 4).[7] They are also higher than with either of the predicting variables alone (rows 1 and 2) or with HM/SE or PGDI/SE (rows 5 and 6). The difference between the correlations obtained with the predicted criterion scores and those with HM/R reached significance when the probability values in both samples are combined. The increase in correlation over HM/SE is not significant.

Prediction 2: A positive correlation between HM/(HM + PGDI) and the criterion (row 4). The correlation is significant in Sample B but not in Sample A. The combined probability is at the .02 level.

Prediction 3a: HM/SE is more positively related to the criterion than HM/R (row 5 greater than row 1). The differences were in the expected direction but small and not significant.

Prediction 3b: PGDI/SE is more negatively related to the criterion than PGDI/R (row 6 greater than row 2). Again the differences are in the expected direction but not significant.

Prediction 3c: A higher relationship to the criterion is provided by a multiple regression type equation employing HM/SE and PGDI/SE than one employing HM/R and PGDI/R (row 7 greater than row 3). Using HM/SE and PGDI/SE in place of HM/R and PGDI/R in the cross-sample multiple regression equations did not result in any particular improvement. However, in comparison with the results stated in prediction 1c, a significant increase in the correlation with the criterion is found with this cross-sample multiple-regression equation employing SE over HM/SE alone (row 7 is greater than row 5).

Relationship of Specific Categories to the Criterion

The correlation coefficients between separate content categories on the Rorschach and ratings of aggressive behavior are presented in Table II. The highest relationships were found with mutilation content which reaches the .01 level of confidence in Sample B and the .10 level in Sample A.

Guilt, Depression, and Inadequacy content combined is also significant in Sample B but gives results in the opposite direction to the prediction in Sample A which increases the combined probability to a level which is no longer significant. The combined probability levels for Hostile content and Passive or Friendly content were, respectively, .12 and .08. The relationship shown by Depressive content are opposite in direction to what was expected but the correlations do not approach significance.

Reliability

Criterion Measures. The reliability coefficients of the criterion measure is presented in Table III. Two forms

TABLE II—Correlation Coefficients Between Separate Content Sub-categories and Average Aggression Ratings

Rorschach Content Sub-category	Sample A (N=15)	Sample B (N=23)
Hostile Content		
Tot. SE	.13	.28*
Mutilation Content		
Tot. SE	.42*	.51***
Passive Content		
Tot. SE	—.40*	—.14
Inadequacy Content		
Tot. SE	.12	—.13
Guilt, Depression, Inadequacy Content		
Tot. SE	.16	—.37**
Hostile Content		
Responses	.41*	.49***
Passive Content		
Responses	.33	—.36**
Inadequacy Content		
Responses	—.42*	—.08
Guilt Content (Presence or Absence)	.10	—.18
Depression Content (Presence or Absence)	.18	.19

* Probability of a chance difference from zero correlation = .05 — .10
** Probability of a chance difference from zero correlation = .01 — .05
*** Probability of a chance difference from zero correlation = .01 or less

of reliability are presented: Intra-subject and Inter-rater. The first of these, Intra-subject, is the split-half reliability of the ratings of each subject and therefore represents the estimated stability of the mean criterion score, obtained over the number of sessions indicated. These reliability coefficients are .95 in Sample A and .91 in Sample B.

Inter-rater agreement, which involved comparison of the mean ratings over all the sessions observed in common, was quite high (r = .89) between the two psychologists in Sample A. It was possible to get independent ratings by one of the two psychologists during only four of the sessions in Sample B, which means that the inter-rater reliability data in this sample is not complete. Agreement between psychiatrist and psychologist observing Sample B (r = .35) was not nearly as good as between the two psychologists observing Sample A. This difference in level of agreement was highly significant (beyond the .001 level of confidence) and suggests that either a systematic difference, because of a somewhat different frame of reference, or greater random error has entered into the ratings made by the psychiatrists. However, the random error could not have been very great in the overall sample in view of the high split-half reliability (intra-subject, above) of the psychiatrists' data.

Reliability of Rorschach Scoring

Coefficients of reliability for the various Rorschach variables and their combinations are presented in Table III. In each sample, there was one extremely long record. Scores of these subjects which were uncorrected for R were well separated from the rest of the scatter plot and, hence, dominated the covariance, spuriously raising correlations. Consequently the reliability coefficients of HM, PGDI, SE, and R, in Table IV, were computed without these two subjects. The full sample coefficients are given in a footnote. Of the variables actually used in prediction, the best reliabilities are obtained with HM/R and PGDI/R. Introduction of total self-expression (SE) reduces the reliability. It might be noted that, in spite

TABLE III—Reliability Coefficients for the Rorschach Scoring

Rorschach Variable	Sample A (15 Subjects)	Sample B (23 Subjects)
HM	.984*	.907*
PGDI	.919*	.881*
R	.980*	.998*
SE	.934*	.983*
HM/R	.94	.86
PGDI/R	.84	.80
HM/SE	.70	.78
PGRI/SE	.53	.84
HM/(HM+PGDI)	.69	.80

* In each of the two samples there was one record of such great length as to be well separated from the rest of the correlational scatter plot. Consequently the reliability of the number of responses (R), or of any score highly related to R (i.e., HM, PGDI, and SE), will be spuriously inflated by the tremendous contribution of this single record to the co-variance in each sample. The reliabilities presented in the body of the table are, therefore, those calculated without these two deviant records. Including the records raises the reliability figures to .990, .996, .9992, and .996 in Sample A, and to .956, .936, .9991, .995 in Sample B, for HM, PGDI, R, and SE, respectively, in each sample.

of their greater unreliability, HM/SE and PGDI/SE are more highly correlated with the criterion than HM/R and PGDI/R. Correction for the attenuation due to unreliability in scoring would increase the margin of this difference even a little further. The correlation between HM/(HM + PGDI) and the criterion would also be improved in relation to the correlations obtained using either HM/R or PGDI/R.

CONFIRMATION OF HYPOTHESES

1. *HM and PGDI content are independently related to aggressive behavior.*

With the limitations in generalization imposed by the nature of the criterion situation and of the sample, an independent relationship between the proportion of HM content in the Rorschach and differences in at least one mode of expression of aggression has been demonstrated. An independent relationship of a lower order, between aggression and the proportion of PGDI content, is also suggested by the near significant results, and this content seems to contribute to prediction when a multiple regression equation is used.

2. *The relationship between HM and PGDI content is the crucial factor in relating this content to aggressive behavior and the ratio of HM content to HM + PGDI content is related to the likelihood that an aggressive impulse, when aroused, will be expressed rather than inhibited.*

The evidence indicates that the particular function used in this study to inter-relate the two types of content is actually a poorer predicter than an appropriate multiple regression type of equation. However, this particular function, HM/(HM + PGDI), did predict significantly better than chance.

3. *The proportions of HM and PGDI contents in the record are more accurately represented in relation to SE than in relation to R.*

None of the predictions related to this hypothesis were confirmed by a significant difference although, with only one exception, the correlations obtained using SE exceeded those obtained using R as a base.

DISCUSSION

The results of this study further indicate and define the potential validity of the content of Rorschach responses as a source of information about the personality of the Rorschach examinee. The correlations obtained in this study are too small to be used for practical individual prediction of behavior. However, the

possibility of using projective material, and Rorschach content in particular, to make explicit predictions about overt behavior is suggested.

Two questions are posed by the results. First, can the discrepancies, between the results of studies by Elizur (6) and by Sanders and Cleveland (18) and the results of other studies (including the present one), be accounted for? Secondly, what are the limitations to the generalization of the present results? For answers to these questions, the scoring and classification of Rorschach content, the criteria employed, and the nature of the subject population must be considered. The discussion of these two questions will not be sharply separated in what follows.

The schema for analysis of aggressive Rorschach content (HM) used in this study, though more extensive, is most similar to that developed by Elizur and used in the studies by Sanders and Cleveland, and Gorlow, Zimet, and Fine (9). In spite of this, the studies by Elizur and by Sanders and Cleveland are the only two studies reporting an inverse relationship between aggressive content and aggressive behavior, in contradiction to the findings of the present study and other previous studies. The study by Gorlow, Zimet, and Fine was generally supportive of a positive relationship between content and behavior. The scale of aggressive content developed by Walker (25), by Finney (8), and by Towbin (23), contain considerable content which would be classified as P, G, D, or I in the present study (e.g., "torn flower," "drunk who fell on his back," "people with dunce caps on," "a rabbit running away," "a man sick with disease," "flea," "decaying," etc.) These differences are sufficiently marked in the case of Finney's scale to account for his comparative lack of results. It also points up the possibility that the PGDI content used in the present study may contain conflicting or invalid items which would account for its lack of relationship to aggressive behavior.

In contrast to previous studies concerned with predicting overt forms of aggression, the criterion in the present study was concerned with a much milder form of aggression and with a purely verbal, as opposed to a physical expression of aggression. Also the situation differed in that the criterion behavior in the previous studies, assault, was invariably discouraged whereas, in the present study, the verbal venting of hostile feelings was permitted, if not encouraged. However, the fact that this study fits in with the majority of findings, particularly when the nature of the subject population is considered (see below), suggests that, with at least certain types of subjects, a positive relationship between aggressive content on the Rorschach and aggressive behavior may hold true with respect to a fairly wide variety of situations and modes of aggressive expression. Despite all the complicated factors determining aggressive behavior in the prison situation, it seems reasonable that individual differences in drive and ability to express aggression will contribute to aggressive behavior in this situation as well as in many other types of situation.

Differences in the type of subjects used may partially explain the striking differences obtained between the results of the present study, that by Storment and Finney (21), by Finney (8), and by Towbin (23), as compared with those obtained by Sanders and Cleveland (18) and predicted by Elizur (6) on the basis of his results. Finney, Storment and Finney, and Towbin found a positive relation between aggressive content and overt aggressive behavior in a group of largely psychotic patients. In this group, in other words, many of the ego (or superego) controls had broken down. On the other hand, the graduate psychology student subjects of Sanders and Cleveland, and the

thirty student volunteers used as subjects by Elizur, could be expected to have predominantly the type of defense-structure in which control is particularly emphasized. The prison inmate population of the present study, while not overtly psychotic, has nonetheless demonstrated certain deficiencies in impulse control by committing the antisocial acts, usually a long history of them, which led to their incarceration. In this respect they may be more like psychotic patients than like the presumably well-controlled college and graduate school students. In the psychotic and prison group, where at least some disturbance in the capacity to control impulse expression can be assumed, one would expect that aggressive drive would achieve overt expression in proportion to its strength. With both a strong need to control and a strong ego, however, one would expect powerful aggressive impulses to arouse still more powerful defensive measures and greater reaction formation. Thus, with poor control one may find a positive correlation between content (which was earlier hypothesized to be an index of underlying drive) and behavior, while with strong control one may find a negative correlation.

A study by Clark (2) lends support to such a formulation. He found that his male student subjects expressed significantly less sexual content in their Thematic Apperception Test productions after being stimulated by pictures of nude women than a group not exposed to the nude pictures. However, under the influence of alcohol, which is known to loosen ego controls, these results were just the reverse. Thus, under alcohol an increase in sexual drive leads to an increase in verbalized sexual fantasy, whereas increased drive results in a decrease in such fantasy without alcohol. If the verbalization of sexual fantasy is equated with overt aggressive behavior, on the basis that both are expressions of culturally disapproved drives, the implications of Clark's study are clear.

The general socio-cultural background of the prison population may also be important in explaining the high relationship found between aggressive content alone and aggressive behavior in this study. Many of these men were raised in an environment which not only condones overt aggression, but in many instances encourages it (3). Consequently, aggressive behavior is probably much more ego-syntonic in this group as a whole than in most other groups and the men behave more nearly in accordance with the strength of their aggressive drives. On the other hand, passivity would probably be much less acceptable in this group than in most other groups. Thus, it might be expected that those inmates who had strong aggressive drives would tend to express them, but that many of those without particularly strong basic aggressiveness might behave in a more assertive, aggressive manner as a reaction against the passive trends in their personality. This leads to the prediction that there would be a greater relationship between the Rorschach Scores and the criterion among those subjects who gave a lot of aggressive content (i.e., are basically aggressive) than in those who have relatively little aggressive content (i.e. are not basically aggressive). Splitting at the median of HM/SE provided four groups in the two samples, two with high aggressive content and two with low aggressive content. The high aggression groups in Samples A and B, respectively, produced correlations of .24 and .64 between HM/SE and the criterion. The corresponding correlations in the low aggression groups were .13 and .43. These differences between high and low Rorschach aggressive content groups are not significant. Also, since the relative stability, and hence reliability, of the larger proportions will be greater, a better relationship with the criterion would be expected for that reason

alone. However, it was further noted that, of the seven cases in the two samples in which the standard score of the criterion was more than one standard deviation greater than the standard score for HM/SE (i.e., behavioral aggressiveness exceeded Rorschach aggressiveness), six were at, or above, the median in the proportion of content scored in the passive category classification. These results, while by no means conclusive, are mildly suggestive. These considerations underline the importance of interpreting projective test data in relation to the socio-economic background of the subject (see also 13).

SUMMARY

The present study sought to relate the content of the Rorschach response to overt aggressive verbal behavior through an approach utilizing both aggressive content and content believed to reflect inhibitory forces in the personality. In addition to its content significance, each response was scored for the degree of self-expression implicit in it. Only records which were sufficiently productive to attain a self-expression score of 10 were used. The subjects were 38 state prison inmates and the criterion was behavior in therapeutic discussion groups. The following results were obtained:

1. The proportion of aggressive content, particularly that dealing with Mutilation content, was significantly positively related to aggressive behavior although the relationships were not high enough for precise individual prediction.
2. The proportion of inhibitory content was inversely related to aggressive behavior but not to a significant degree (the .15 level).
3. A multiple regression type of equation, developed independently of the sample to which it was applied, predicted significantly better than an equation which would reflect an interaction between aggressive and inhibitory content.
4. Improvement in prediction, when the total self-expression score was substituted for the number of responses in calculating the proportion of all content which was aggressive or inhibitory, failed to reach statistical significance. However, inhibitory content was significantly inversely related to aggressive behavior when total self-expression was used in obtaining the proportion.

REFERENCES

1. Brosin, H. E. Contributions of psychoanalysis to the study of the psychoses. In F. Alexander and Helen Ross (Ed.), *Dynamic psychiatry*. Chicago: University of Chicago Press, 1952. Pp. 285-306.

2. Clark, R. A. The projective measurement of experimentally induced levels of sexual motivation. *J. exp. Psychol.*, 1952, *44*, 391-399.

3. Davis, A. Socialization and adolescent personality. *Adolescence, Forty-Third Yearbook*, Part I. National Society for Study of Education. Chicago: University of Chicago Press, 1944.

4. De Voss, G. A quantitative approach to affective symbolism in Rorschach responses. *J. proj. Tech.*, 1952, *16*, 133-150.

5. Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H.; and Sears, R. R. *Frustration and aggression*. New Haven: Yale University Press, 1939.

6. Elizur, A. Content analysis of the Rorschach with regard to anxiety and hostility. *J. proj. Tech.*, 1949, *13*, 247-284.

7. Fenichel, O. *The psychoanalytic theory of neurosis*. New York: W. W. Norton, 1945.

8. Finney, B. C. Rorschach test correlates of assaultive behavior. *J. proj. Tech.*, 1955, *19*, 6-16.

9. Gorlow, L., Zimet, C., and Fine, H. J. The validity of anxiety and hostility Rorschach content scores among adolescents. *J. consult. Psychol.*, 1952, *16*, 73-75.

10. Hertz, Marguerite R. *Frequency tables for scoring responses to the Rorschach inkblot test*. (3rd Ed.) Cleveland: The Press of Western Reserve University, 1951.

11. Mosteller, F., and Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. *Theory and method*. Cambridge: Addison-Wesley, 1954. Pp. 289-334.

12. Mussen, P. H., and Naylor, H. K. The relationships between overt and fantasy aggression. *J. abnorm. soc. Psychol.*, 1954, *49*, 235-240.

13. Phillips, L., and Smith, J. G. *Rorschach*

interpretation: advanced technique. New York: Grune & Stratton, 1953.

14. Pittluck, P. The relation between aggressive fantasy and overt behavior. Unpublished doctor's dissertation, Yale University, 1950.

15. Quenouille, M. H. *Associated measurements*. London: Butterworth, 1952.

16. Rapaport, D.; Schafer, R.; and Gill, M. *Diagnostic psychological testing*. Vol. II. Chicago: Year Book Publishers, 1946.

17. Redl, F., and Wineman, D. *Children who hate*. Glencoe: The Free Press, 1951.

18. Sanders, R., and Cleveland, S. E. The relationship between examiner personality variables and subjects' Rorschach scores. *J. proj. Tech.*, 1953, *17*, 34-50.

19. Sanford, R. N., et al. Physique, personality and scholarship. *Monographs of the Society for Research in Child Development*, Vol. VIII, 1943, No. 1 (Serial number 34). Washington, D. C.: Society for Research in Child Development, National Research Council.

20. Schafer, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954.

21. Storment, C. I., and Finney, B. C. Projection and behavior; a Rorschach study of assaultive mental hospital patients. *J. proj. Tech.*, 1953, *17*, 349-360.

22. Symonds, P. M. *Adolescent fantasy*. New York: Columbia University Press, 1949.

23. Towbin, A. P. Hostility in Rorschach content and overt aggressive behavior. Unpublished doctor's dissertation, Yale University, 1955.

24. Walker, Helen H., and Lev, J. *Statistical inference*. New York: Henry Holt, 1953.

25. Walker, R. G. A comparison of clinical manifestations of hostility with Rorschach and MAPS test performances. *J. proj. Tech.*, 1951, *15*, 444-460.

26. Whitehorn, J. C. Psychodynamic approach to the study of psychoses. In F. Alexander and Helen Ross (Ed.), *Dynamic psychiatry*. Chicago: University of Chicago Press, 1952. Pp. 255-284.

27. Zubin, J. Failures of the Rorschach technique. *J. proj. Tech.*, 1954, *18*, 303-315.

Received October 18, 1956

Rorschach Findings in a Group of Peptic Ulcer Patients and Two Control Groups [1]

IRVING RAIFMAN
United States Naval Hospital, National Naval Medical Center, Bethesda, Maryland

As part of an investigation of the personality factors of dependency and over-compensatory goal striving behavior associated with the development of peptic ulcer the Rorschach was administered to fifteen peptic ulcer patients, fifteen psychoneurotic patients, and a like number of subjects classified as normal. At the time the study was undertaken, the temper of psychological studies with the Rorschach test was predominantly statistical in approach. The investigator, adhering to the times, attempted to isolate ulcer patients from the two control groups by analyzing the data in statistical terms. A few clinical departures were attempted in an effort to study content material and analyze the quality of M and FM responses. However, a case by case study could not be done, and the qualitative data was also subjected to statistical treatment. To overcome the loss of what Piotrowski calls the "interdependence of components" (7, p. 80), the author submitted the Rorschach protocols along with other test data to two judges who rated each case according to a devised rating list. This aspect of the study is not pertinent to the present paper and will be discussed in a future paper.

The present paper is meant to be a critical analysis of attempts made to treat the Rorschach statistically and graphically without full regard for the essential "rule" of Rorschach interpretation, that every symbol has a conditional or variable meaning and is dependent upon other symbols for a complete interpretation.

Though there have been many studies of the psychological factors associated with the development of peptic ulcer, the literature does not indicate that the Rorschach was ever used prior to the report of Ruesch and his coworkers (8) in 1948. Since then, Brown, et. al. (3), Hecht (4), Osborne and Sanders (6), Brennan (2), have all used the Rorschach in the study of peptic ulcer. A brief review of their work indicates that few group studies have proved the use of the Rorschach test to be of value in differentiating one group from another.

It was the author's intention to investigate the "primary conflict" situation of ulcer patients as described by Alexander and others (1). Briefly, Alexander hypothesized that ulcer patients show intense receptive and acquisitive wishes against which they fight internally because the wishes are connected with extreme conflict in the form of guilt and feelings of inferiority. Tendencies toward overactivity, efficiency, and ambitious effort are compensations to mask the strong dependent wishes. Three hypotheses were put to the test, namely,

a) Peptic ulcer patients express significantly more dependency needs than normal and psychoneurotic patients.

b) The dependency needs of peptic ulcer patients are significantly less acceptable to them than are the dependency needs of the control groups.

[1] This paper is based on a doctoral dissertation submitted to the department of psychology, New York University in 1951. The study was discussed at the meetings of the A.P.A. in New York City, September 1954. Grateful acknowledgment is due Dr. B. Tomlinson and other members of the supervisory committee, Dr. R. Morrow and his staff at the Bronx VA hospital, New York, N. Y. for their help and cooperation. The author's views are his own and in no way reflect the opinion of the Navy Department or the VA.

c) Peptic ulcer patients substitute a facade of overcompensatory goal striving activity to mask their feelings of inadequacy associated with dependency needs whereas the control groups do not.

Method

The subjects were selected from the General, Medical, and Surgical wards of the Kingsbridge VA hospital in New York City. They were chosen from the white population and were native born citizens of the United States with the ability to read and write English. Their ages ranged from twenty-two through forty-five years. The selection of patients was not limited to age, education, religion, marital status or an occupational group. Still, there were no significant differences with respect to these factors among the three groups. The experimental group consisted of fifteen peptic ulcer patients, whose ulcers were demonstrated by X-ray data and diagnosed as such by the ward physicians. None had a history of central nervous system trauma or disease. The two control groups were non-ulcer patients. One control group of fifteen patients was selected with the understanding that the patients had never had a peptic ulcer or a history of psychoneurosis. After each subject had satisfied the criteria established in the Cornell Selectee Index (9) this group was considered the "normal." The group was selected to equalize the effects hospitalization may have had on the experimental group and as a base line of reference. Whereas the normal group was comprised of patients with minor surgical problems, the other control group was comprised of fifteen patients selected because of a history of psychoneurosis without organic lesion. The neurotics were chosen to determine whether or not the dynamics under investigation were also typical of them inasmuch as numerous authors regard the ulcer patient as psychoneurotic.

The Rorschach was given each of the forty-five subjects. The Klopfer and Kelley (5) scoring technique was followed. Tabulations were made according to 35 variables established by the Klopfer method of scoring. In addition calculations were made according to the number of patients in each group for whom the various Rorschach signs and ratios were present.

There were several signs and variables introduced by the author which require explanation. In Table I *All Shading* includes Fc, Fk, KF, and FK responses. *All Objects* include Obj., AObj. The *Content Denoting Ego Strength* refers to the number of responses which suggest power, aggression, greatness, status, striving, masculinity, i.e., such associations as crowns, eagles, lobsters, bats, fire, dogs, bears, monsters, emblems, wolves, men, or any percept seen as large and strong. On the other hand, the *Content Denoting Passivity* refers to the number of responses suggesting weakness, dependence, femininity, i.e., responses such as butterflies, cows, babies, cubs, moths, flowers, scenery, vases, women, anatomy, or percepts seen as small and weak. Extensor movement and flexor movement concepts refer to assertive and passive projections of phantasy respectively.

Results and Discussion

The tables tell the story of the fruitless effort to evaluate group data with the purpose of distinguishing one nosological group from another by a quantitative analysis. One hundred five comparisons were made with respect to the 35 quantifiable signs shown in Table I. The groups were found to be significantly different from one another in only five of the quantifiable signs. A total of seven significant differences was obtained with respect to these five signs (see Table II). This total is too small a number of significant differences to be important since it could occur by chance alone. In any one hundred "t"

TABLE I—Means (M) and Standard Deviations (SD) for the Quantifiable Rorschach Test Data for the Ulcer, Normal, and Psychoneurotic Groups

Sign	Ulcer		Normal		Neurotic	
	M	SD	M	SD	M	SD
R	16.80	7.99	23.80	7.99	22.07	8.08
W, WS	5.33	2.67	4.93	4.33	4.33	2.37
D, DS	10.60	3.35	15.53	8.35	14.87	5.26*
d, dd, etc	1.07	1.89	2.93	3.69	3.07	4.37
All S	.93	2.05	1.13	2.05	.80	.45
F	7.53	5.46	12.00	8.25	10.47	5.29
M	2.33	2.21	2.40	1.93	2.40	2.21
FM	3.60	1.36	4.47	2.19	3.60	1.99
m	.73	1.24	.40	.71	.53	.62
All Shading	1.40	1.25	1.93	1.57	2.80	2.17*
FC	.73	.93	1.13	1.41	1.20	1.17
CF	1.33	2.02	1.60	1.45	1.67	1.57
C	.33	.59	.20	.54	.20	.40
H	1.33	1.25	1.53	1.52	2.00	1.71
A	7.20	2.96	8.13	3.44	6.93	2.74
Hd	1.47	3.14	2.47	2.86	2.20	2.17
Ad	1.53	1.25	3.93	4.07	2.80	1.97*
All Objects	2.13	2.06	2.73	3.11	3.73	2.79
At, X-ray	.93	1.06	1.80	1.87	1.47	1.31
Nat., Bot., etc	1.53	1.91	2.20	1.80	2.00	1.46
P	5.60	1.45	5.13	1.36	5.67	1.47
F%	44.51	16.43	47.13	12.04	45.98	14.09
F+%	90.25	10.09	90.41	8.74	92.78	9.10
% VIII-X	33.81	8.20	38.27	6.94	32.86	15.33
Av. r.t. Achr'tic	17.20	9.15	16.97	10.55	19.20	11.70
Av. r.t. Chr'tic	19.83	4.94	16.72	9.19	25.28	16.49
A%	58.89	20.05	51.57	11.14	47.15	19.86
H%	13.90	11.09	16.33	9.20	18.31	13.04
W%	32.55	14.52	25.21	18.05	22.93	18.28
Content denoting Ego Strength	6.47	2.99	9.87	5.13	8.81	4.01*
Content denoting Passivity	9.60	6.30	12.20	6.50	11.60	4.20
Extensor M	.93	1.24	1.07	1.00	.93	1.18
Flexor M	1.27	1.34	1.07	1.00	1.47	1.26
Extensor FM	2.27	1.44	2.07	1.39	2.13	1.46
Flexor FM	1.13	.89	1.67	1.13	.67	.85*

* There are significant differences among the groups for these quantifiable signs.

TABLE II—t Test Values for those Quantifiable Rorschach Signs Showing Significant Differences Among the Three Groups

Sign	Ulcer-Normal	Ulcer-Neurotic	Normal-Neurotic
D, DS	2.08*	2.54**	.25
All Shading	.98	2.12*	1.22
Ad	2.14*	2.05*	.98
Content denoting Ego Strength	2.16*	1.77	.62
Flexor FM	1.42	1.40	2.73***

* Significant at the .05 level
** Significant at the .02 level
*** Significant at the .01 level

tests made, five could reach the .05 level, two the .02 level, and one the .01 level. What then have we achieved? Two of the seven differences would have meaning.

Let us examine the sign D, DS for example. The ulcer group gave fewer such responses than did the normal and the psychoneurotic groups. According to standard interpretation this would signify that the ulcer group uses less common sense in the handling of daily routines, a finding which would fit in with what we

know of persons suffering from ulcers. However, this is not a significant figure when the number of D, DS responses is interpreted in terms of the total number of responses given by each group. The percentage of D responses given by all groups was quite similar, for the ulcer group 63, for the normal group 65, and for the psychoneurotic group 67. Thus the fewer absolute number of D, DS responses given by the ulcer group cannot be

TABLE III—The Number of Patients in Each Group for Whom the Various Rorschach Signs were Present

Sign	Ulcer	Normal	Neurotic
Card Rejection	6	6	3
Blood	3	4	3
Fire, Flame, Explosion	6	1	4
Map	4	4	4
Cloud	2	4	3
Food	4	4	3
Combinatory W	9	8	5
W:M ratio			
W>2M	9	7	5
W<2M	4	6	8
W=2M	2	2	2
M: sum C ratio			
Coarcted	3	0	1
Ambiequal	2	4	2
Predominantly M	6	5	5
Predominantly C	4	6	7
H+Hd:A+Ad			
A+Ad>2 H+Hd	10	14	9
A+Ad=2 H+Hd	3	0	1
A+Ad<2 H+Hd	2	1	5
W% greater than 30%	11	5	2
A% = 50% or greater	9	11	7
H% = 20% or greater	5	7	8
Extensor M:Flexor M			
Extensor>Flexor	2	1	3
Extensor<Flexor	4	2	0
Extensor=Flexor	2	3	5
Coarcted	7	9	7
Extensor FM:Flexor FM			
Extensor>Flexor	6	4	6
Extensor<Flexor	0	2	0
Extensor=Flexor	4	6	4
Coarcted	5	3	5
Passive Percepts:Ego Strength Percepts			
Passive>Ego Strength	10	8	10
Passive<Ego Strength	1	3	1
Passive=Ego Strength	4	4	4

TABLE IV—P Values Obtained in the Evaluation of Differences in the Incidence of Various Signs

Sign	X^2	df	P*
Card Rejection	1.80	2	.50>P>.30
Combinatory W	2.31	2	.50>P>.30
W% greater than 30%	11.67	2	.01>P>.001
A%=50% or greater	2.22	2	.50>P>.30
H%=20% or greater	1.25	2	.70>P>.50

* Chi square cannot properly be applied whenever the hypothetical frequency for any cell in the matrix becomes too small, i.e., less than five. This limitation in the use of Chi square prevented the calculation of significance tests in certain cases.

TABLE V—P Value Obtained in the Evaluation of Differences in the Incidence of Various Rorschach Ratios

Ratios Compared	X^2	df	P
W:M			
W>2M and W<2M	2.48	2	.30>P>.20
W>2M and Other W:M	2.14	2	.50>P>.30
M:C			
Predom M and Predom C	.78	2	.70>P>.50
Predom M and Other M: sum C	.20	2	.95>P>.90
Predom C and Other M: sum C	1.32	2	.70>P>.50
Extensor M:Flexor M			
Coarcted and Other M types	.71	2	.70=P
Extensor FM: Flexor FM			
Extensor>Flexor and Other FM	.77	2	.70>P>.50
Passive: Ego Strength Percepts			
Passive>Ego Strength and Other types	.76	2	.70>P>.50

construed as really meaningful.

There was a significant difference between the ulcer and the psychoneurotic groups with respect to the use of shading. The psychoneurotics gave twice as many shading responses as did the ulcer group. An abundance of shading concepts is considered to be indicative of anxiety, uncertainty, of a feeling of being exposed to danger, of considering the environment to be hostile, and of doubt concerned with the most suitable method of restoring security. This difference among the two groups is quite plausible if it can be construed as meaningful. The attempt to see a close kinship between the ulcer group and the psychoneurotic group, which some authors have made, is not supported. One would expect that ulcer patients convert their anxiety so that it is not obvious in the same way that psychoneurotics manifest their feelings.

In Table I, Ad stands out as an important variable because the groups were found to be dissimilar on this score. The relative lack of Ad in the ulcer group would suggest that these subjects were less critical toward the form qualities of the blots than were the other two groups. If one were to generalize, it would be possible to say that the ulcer group was less critical because it was too inclined toward making generalizations. However, this would be stretching the point too far. At best one can say very little about the meaning of the "significant" Ad finding.

One would expect the ulcer patients to project their desires for power, prestige, greatness, achievement, strength and aggressiveness onto the blots. The results of the mean number of such percepts called *Content denoting Ego Strength* indicate otherwise. As can be seen in Table I, the normal group gave significantly more such percepts than the ulcer group. To make this finding meaningful, one would have to say that the normal subjects were inherently more assertive and have more ego strength than the ulcer patients. If one observes the conscious efforts of the ulcer patients to prove their worth this result introduces a note of confusion. Why don't they show more assertiveness in their Rorschach concepts? Is the content material of an unconscious variety? One would have to think so if the finding is to be interpreted logically and in a meaningful way. Though one would expect the ulcer patients to project more of their need for independence and assertiveness, they manifested less of it in comparison with normal subjects. This would imply that they are motivated in a compensatory way, so that while they appear to be strivers, striving is not an intrinsic, basic characteristic of their personality structure.

The psychoneurotic group offered significantly fewer passive, dependent

FM responses than did the normal group (as noted in Table II), but was not different from the ulcer group. It is too isolated a finding to fit into any particular scheme even in terms of the dynamics involved in the psychoneurotic problem. An effort to explain this finding in dynamic terms as to its meaning in relation to the three groups under investigation yields nothing but jargon phrases and tautological thinking, unless one challenges the currently accepted interpretation of FM responses.

Of all the signs listed in Table III, one significant finding is obtained, namely, that there is a greater number of ulcer patients whose W% exceeds 30%. This isolated finding is in harmony with the expected results. Ulcer patients are more concerned with demonstrating ambition, drive and achievement.

The results obtained when quantified Rorschach scoring symbols are used to compare one nosological group with other groups are discouraging. Few significant differences emerge that are useful and related. It would seem, consequently, that the Rorschach method is a fruitless technique when used quantitatively. Quantification eliminates and/or destroys the dynamic picture of personal forces, defenses, sequences and balances that may be observed when the Rorschach is studied as a gestalt.

Summary

An attempt has been made to show that the Rorschach technique is not a quantifiable tool in the usual sense, and that when it is so utilized to differentiate specific character structures the positive results are minimal. Several hypotheses based on the description by Alexander and others of the personality pattern of ulcer patients were considered. A group of peptic ulcer patients and two control groups were used. Of one hundred and five comparisons made on the basis of various Rorschach signs only seven significant differences were obtained. Of sixteen other ratios tested by the chi-square technique, one was significant. Some of the findings in part uphold the hypotheses under investigation, but, as with the earlier Rorschach studies cited, it would be necessary to force many meaningful interpretations or to resort to tautological reasoning and/or challenge the very essence of the Rorschach test if the quantitative technique is used.

References

1. Alexander, F., and French, T. *Studies in psychosomatic medicine.* New York: Ronald Press, 1948.

2. Brennan, M. A personality study of patients suffering from peptic ulcer. Unpublished doctor's dissertation, Fordham University, New York, 1950.

3. Brown, M., Bresnahan, T. J., Chalke, F. C. R., Peters, Barbara, Poser, E. G., and Tougas, R. V. Personality factors in duodenal ulcer; a Rorschach study. *J. psychosom. Med.*, 1950, *12*, 1-5.

4. Hecht, I. The differentiation of certain psychosomatic groups in terms of psychometric patterns. Unpublished doctor's dissertation, New York University, New York, 1949.

5. Klopfer, B., and Kelley, D. *The Rorschach technique.* New York: World Book, 1942.

6. Osborne R. T., and Sanders, W. B. Rorschach characteristics of duodenal ulcer patients. *J. clin. Psychol.*, 1950, *3*, 258-262.

7. Piotrowski, Z. A Rorschach compendium. *Psychiat. quart.*, 1947, *21*, 79-101.

8. Ruesch, J., et al. *Duodenal ulcer; a sociopsychological study of Naval enlisted personnel and civilians.* Berkeley: University of California Press, 1948.

9. Weider, A., Wolff, H. G., Brodman, K., Mittelmann, B., and Wechsler, D. *Manual for the Cornell Index.* New York: Psychological Corporation, 1948.

Received March 29, 1957

A Comparison of Several Measures of Psychosexual Disturbance[1,2]

ALEXANDER TOLOR
USAF Hospital, Parks Air Force Base, California

The purpose of this study is to report on the performance of a large group of patients on three different measures of psychosexual confusion. The techniques selected in this investigation are all methods which many clinicians regard as providing important information concerning the adequacy in the identification with and acceptance of one's sexual role. The projective devices which were employed are the human figure drawing technique and the M. R. Harrower Psychodiagnostic Inkblots. The choice of the Harrower blots in preference to the standard Rorschach series was prompted by the fact that we intended to save the latter device without prior contamination for actual diagnostic purposes. Inasmuch as the Harrower blots are purported to represent a series of stimuli which are parallel to the Rorschach blots, it was thought that their use in this research was justified and that conclusions relating to one series of blots would probably be equally germane to the other. In addition to these projective measures, an objective scale developed by Gough (3), and allegedly capable of assessing masculinity or femininity of interests, was also administered.

It has been a common practice of clinical psychologists to assign interpretative significance to the sex of the figure drawn first on the Draw-a-Person Test. The underlying hypothesis is that an individual who draws the opposite sex first is suspected of having a serious disturbance in psychosexual identification and one who draws a figure of his own sex first does not. However, an extensive review of the relevant literature by Brown and Tolor (1) led to the conclusion that sweeping interpretative generalizations which are based solely on this particular aspect of the figure drawing technique seem unwarranted. Similarly, it is a general practice among clinicians to assign dynamic significance to the sex which the patient ascribes to the people which he perceives in the ambiguous stimuli of inkblots, particularly the popular figures on card III. The male patient, in this theoretical formulation, is thought to manifest signs of an inadequate or confused sexual identification if he perceives the popular figures on card III as something other than two male figures (6, 10, 12). The question arises as to whether these two measures of sexual disturbance are highly related to one another, i.e., whether a tendency to draw a figure of the opposite sex is in fact frequently associated with the tendency to make an opposite-sexed identification in the inkblots. Another related problem that seems worthy of attention is the ascertaining of whether these two projective measures of psychosexual orientation are in any way related to the results obtained from a third but objective measure of masculine-feminine interests, such as the Gough scale.

METHOD

The subjects comprising the experimental population were 122 consecutive U. S. Air Force male recruits in various stages of their basic training. Each of these airmen had been referred to the Mental Hygiene Clinic at Parks Air Force Base in California for evaluation of his adjustment problems. When an analysis of the

[1] The opinions and conclusions expressed by the author do not necessarily represent those of the Department of the Air Force.

[2] This paper was presented at the 1957 meeting of the Western Psychological Association.

final diagnoses of these patients is performed, one finds the following psychiatric conditions represented in the sample: immaturity with symptomatic habit reaction, manifested by a history of enuresis[3] extending at least to the age of 15 years—38; emotional instability reaction—48; inadequate personality—13; no specific psychiatric disease—10; schizoid personality —6; anti-social personality—2; asocial personality—2; sexual deviate, overt homosexuality—2; and schizophrenic reaction, not elsewhere classified—1. The age range of this sample is 16 to 25 years, the median age being 18. As for the educational level attained, the range is 7 to 16 years, with a median of 11 years.

These patients had been referred to the Mental Hygiene Clinic by various sources on the base including other medical installations (flight surgeon's office, dispensaries, and medical officer on duty at night or over weekends), organization commanders, chaplains, and stockade officials. Prior to interviewing by either psychiatrist, psychologist, or psychiatric social worker,[4] each of the patients was seen by one of the enlisted technicians on the clinic staff who routinely administered the figure drawing test, card III of the Harrower Psychodiagnostic Inkblots, and the Gough Femininity scale in individual testing.

On the figure drawing test patients were initially requested to draw a person and upon completion, to make a person of the opposite sex on a new sheet of paper. On the inkblot test the instructions accompanying card III were the same as are generally considered standard on the Rorschach. All responses were recorded verbatim. In those instances in which patients spontaneously identified the lateral large areas of the blot as people of a specific sex, no further inquiry was necessary. However, when no reference was made to the sex of the figures, further probing was attempted. Whenever a patient failed spontaneously even to identify the lateral areas as people, the examiner informed the patient that such a response is frequently given and determined whether the patient would acknowledge two human figures. In case a human concept was finally conceded, an attempt was then made to have the patient identify the sex.

[3] A number of other patients presented a history of enuresis but were assigned different diagnoses which were thought to reflect more accurately the overall symptomatology.

[4] Appreciation is extended to Dr. Simon Dorfman, psychiatrist and officer in charge of the clinic, and to 2nd Lieutenant John E. Davis, psychiatric social worker, for their cooperation in this project.

Results

An analysis of the frequency with which like-sexed and opposite-sexed figures are drawn indicates that of the basic trainees tested 87% made figures of the same sex first as compared to 13% drawing a figure of the opposite sex first. Since the absence of clothing in human figure drawings is often interpreted as suggesting the presence of a pathological self-preoccupation and "body narcissism" (2, 4, 7, 8), the data on the number of patients drawing nude and clothed figures were also obtained. It was noted that 75% of the patients drew both figures clothed. Table I presents data indicating that there is little difference in the frequency with which nude and clothed figures are portrayed by patients drawing the same sex first as compared to those drawing the opposite sex first.

Table I — Frequency With Which Nude Figures Are Made by Patients Drawing the Same and the Opposite Sex First.

Figures Drawn	Sex Drawn First	
	Same (N = 106)	Opposite (N = 16)
	(In Percentages)	
Both Clothed	76	69
Both Nude	17	25
Only Female Nude	4	6
Only Male Nude	3	0

With respect to card III of the Harrower inkblots, 70% of our sample of trainees spontaneously offered some type of human response to the usual area of this card. The other 30% did not initially perceive a human figure in this blot, and 7% failed to acknowledge the possibility of a human concept even when it was specifically suggested in the testing of the limits phase of the presentation.

As for the sex of the figures visualized, 40% of the sample regarded them as being two male figures, 24% thought they were both females, 18% considered one of the pair as male and the other as female, and another 18% was unable to arrive at any specific sexual differentiation whatsoever even when specifically urged to do so. Thus, a total of 60% of our sample interpreted the popular human figures as something other than two males. If one were to accept the projective test postulate that individuals who do not perceive the popular figures on card III as being of the same sex as themselves have serious problems relevant to their sexual identification, then a very high proportion of this patient population would be so characterized.

The quality of the human movement response is also believed to be related to the personality variables of ascendancy (masculinity) and submission (femininity) (5, 6, 9, 11). In the present investigation no special effort was made to determine which of the human responses represented associations of the human movement type as no inquiry was conducted for this purpose. However, when only those spontaneous responses are considered in which patients make it explicit without any further prompting that they are using M, 21 self-assertive, 2 compliant, and 1 indecisive response are obtained. Thus, we are faced with the surprising finding that although the majority of patients meets the criterion of sexual confusion in misidentifying the figures on card III, most of these same patients utilize the type of M which is classified as being assertively masculine.

The mean Gough scale score for the sample was 28.9 (S. D. 5.4). In comparison to Gough's male criterion group, in which the reported mean score is 24.8 (S. D. 4.6), a greater degree of psychological femininity is indicated for the present population, the difference being highly significant ($t = 6.8$, $p < .001$). Thus, this scale, in contrast to the figure drawing technique but in line with the Harrower inkblots, suggests a marked degree of psychological femininity in this sample of referred basic trainees.

There is no significant difference in mean Gough scores earned by patients drawing a male figure first and those drawing a female figure first. The respective means are 29.0 (S. D. 5.3) and 28.4 (S. D. 5.4). Similarly, the Gough scale scores fail to disclose any significant differences between patients identifying the figures on the inkblot test as males and those not doing so. The former achieve a mean score of 28.8 (S. D. 4.9), the latter obtain the identical mean of 28.8 (S. D. 5.9). When patients falling at the extremes in the Gough score distribution are compared with respect to the frequency with which they draw the same and the opposite sex first, one finds no significant differences between groups (see Table II). Table III also reveals that there is no significant relationship between those patients falling at the extremes in the Gough score distribution and the sex ascribed to the figures on card III of the Harrower blots. These findings seem to establish quite convincingly the absence of any significant

TABLE II—Femininity Scores and Sex of First Drawing

Gough Score	Figure Drawn First: Male	Female
35 & above (N=23)	21	2
23 & below (N=25)	22	3

relationship between Gough scale scores of femininity and the simple factor of which sex is drawn first, and between Gough scale scores and which sex the patient attributes to his inkblot figures.

TABLE III—Femininity Scores and Sex of Figures on Card III.

	Figures in Inkblot Perceived As	
Gough Score	Both Males	Other
35 & above (N=23)	7	16
23 & below (N=25)	7	18

Table IV presents data on the relationship between the tendency to draw the like or the opposite sex first and the type of sexual identification made of the figures on card III. There is no statistically significant association between these two variables either.

TABLE IV—Relationship Between the Sex of the First Drawing and the Inkblot Figures.

Figure Drawn First	Figures on Card III Identified as Both Males		Other	
	(N)	(%)	(N)	(%)
Male	44	41.5	62	58.5
Female	5	31.2	11	68.8

CONCLUSIONS

Within the limitations of the experimental procedures and sample used, the data obtained indicate that these three measures of adequacy of psychosexual adjustment are entirely unrelated, cannot be used interchangeably, and apparently reflect different psychological functions. As a rule, the clinician therefore is not justified in expecting to find the same male patient drawing a figure of the opposite sex first, identifying the popular figures on card III as female, and scoring high in femininity on the Gough scale. No two of these indicators are significantly related to one another. If for the sake of discussion one were to accept the validity of these instruments insofar as they may indicate acceptance or rejection of sexual role, then it would appear that the observation of which sex is drawn first represents a far less sensitive device than the observation of which sex is ascribed to the inkblot figures. It is even conceivable that the latter is a hyper-sensitive measure of psychopathology, pointing toward sexual ambiguity when no other clinical corroborative evidence exists. In any case, because of the methodology employed here one cannot draw any definitive conclusions concerning the relative efficacy of these three different methods for detecting ambiguity in sexual identification.

SUMMARY

A sample of 122 U. S. Air Force basic trainees, who had been referred to a mental hygiene clinic for evaluation of their adjustment problems, was requested to take the human figure drawing test, to respond to card III of the Harrower Psychodiagnostic Inkblots, and to complete the Gough scale of psychological femininity. The performance of the patients on these three selected measures, which are presumed to be capable of determining level of psychosexual functioning, was subjected to analysis.

Among the findings, the following appear to be the most important: (1) 87% of these patients draw the male figure first; (2) 60% interpret the popular human figures on the inkblot test as something other than two males; (3) the mean Gough scale score for the sample indicates far greater psychological femininity in these patients than in a comparable normal criterion group; (4) there is no significant relationship between any two of the selected measures.

It is concluded that these three measures of adequacy of psychosexual adjustment cannot be used interchangeably as they are entirely unrelated and appear to reflect different psychological functions. The relative validity of each should be further evaluated.

REFERENCES

1. Brown, D. G. & Tolor, A. Human figure drawings as indicators of sexual identification and inversion. Unpublished manuscript, 1956.

2. Buck, J. N. *Administration and interpretation of the H-T-P test.* Los Angeles: Western Psychological Services, 1950.

3. Gough, H. G. Identifying psychological femininity. *Educ. psychol. Measmt.*, 1952, *12*, 427-439.

4. Hammer, E. F. *The H-T-P clinical research manual.* Beverly Hills, Calif.: Western Psychological Services, 1955.

5. Hammer, E. F. & Jacks, I. A study of Rorschach flexor and extensor human movement responses. *J. clin. Psychol.*, 1955, *11*, 63-67.

6. Klopfer, B., Ainsworth, Mary D., Klopfer, W. G. & Holt, R. R. *Developments in the Rorschach technique.* Vol. 1, Yonkers-on-Hudson. New York: World Book Co., 1954.

7. Levy, S. Figure drawing as a projective test. In L. E. Abt & L. Bellak (Eds.), *Projective psychology.* New York: Alfred A. Knopf. 1950.

8. Machover, Karen. Drawing of the human figure: A method of personality investigation. In H. H. Anderson & Gladys L. Anderson (Eds.), *Projective techniques.* New York: Prentice Hall, 1951.

9. Mirim, B. The Rorschach human movement response and role taking behavior. *J. nerv. ment. Dis.*, 1955, *122*, 270-275.

10. Phillips, L. & Smith, J. G. *Rorschach interpretation: advanced technique.* New York: Grune & Stratton, 1953.

11. Piotrowski, Z. A. A Rorschach compendium: revised and enlarged. *Psychiat. Quart.*, 1950, *24*, 549-596.

12. Shafer, R. *Psychoanalytic interpretation in Rorschach testing.* New York: Grune & Stratton, 1954.

Received March 4, 1957.

An Exploration of Relationships between Manifest Anxiety and Selected Psychosexual Areas

RICHARD D. TRENT AND ABRAHAM AMCHIN
New York State Training School for Boys

The concept of anxiety is of primary importance in the investigation and understanding of personality. The variety of theory seems to indicate that anxiety is a painful affective state which is produced by actual or imagined threats to the ego (7, 8, 11, 12, 13, 15). Within this body of theory, several authors (9, 11, 12, 22) have related anxiety to specific areas of threat. However, there is a need for further experimental investigation of these relationships.

The purpose of this study was to explore the relationship between "manifest anxiety" and some specific areas of psychosexual conflict. Manifest anxiety was operationally defined in terms of Ss' performance on a questionnaire (6) purporting to measure anxiety. Specific areas of psychosexual conflicts were limited to Ss' performance on the Blacky Pictures (3). The psychosexual areas studied included: oral eroticism, oral sadism, anal expulsiveness, anal retentiveness, oedipal intensity, masturbation guilt, castration anxiety, positive identification, sibling rivalry, guilt feelings, positive ego ideal, narcissistic love object and anaclitic love object.

In addition to the exploration of relationships between manifest anxiety and specific areas of psychosexual conflict, an attempt was made to answer the following questions: (a) Was there a relationship between the degree of manifest anxiety and the number of areas of problems? (b) What areas of psychosexual conflict were most common to the delinquent Ss? (c) Were there significant differences in psychosexual conflicts between white, Negro and Puerto Rican Ss?

SUBJECTS

The Ss consisted of 57 delinquents at the New York State Training School for Boys at Warwick, a residential institution for adjudicated delinquents. Twenty boys were white, 19 were Negro and 18 were of Puerto Rican origin. Thirty-five Ss were Catholic, 22 were Protestant. These Ss were matched so that there were no significant differences between whites, Negroes, and Puerto Ricans in age and intelligence as measured by the Wechsler-Bellevue. The average age of Ss was 14 years, 2 months with a range from 12 years, 3 months to 16 years, 0 months. The range of intelligence scores was from 65 to 120 with an average of 83. All Ss were from deprived socio-economic backgrounds. The two most common offenses which resulted in the Ss' commitment to the training school were truancy from school and stealing.

PROCEDURES

Anxiety was measured by the Children's Form of the Manifest Anxiety Scale (CMAS). This questionnaire was developed by Castaneda (6) and his associates at the Iowa Child Welfare Research Station and was an adaptation of the Taylor Scale of Manifest Anxiety (10, 23, 24). The CMAS scale has been previously employed by several investigators (16, 25) as a measure of manifest anxiety. The questionnaire consists of 41 anxiety items and 11 lying items (L-Scale) which were designed to tap the Ss' tendency to falsify responses. Subjects responded to each item by checking either *Yes* or *No*.

The reliability coefficient of the CMAS test was determined by correlating odd CMAS items with even items.

The reliability coefficient, corrected by the Spearman-Brown formula, was .86 ($p < .01$) which indicates that the performance of Ss was statistically reliable.

The Blacky Pictures were used to determine the Ss' specific psychosexual problems. Considerable use has been made of this test by various investigators (1, 14, 17, 18, 19, 20). This instrument was developed by Blum (2, 3, 4) and consists of eleven cartoons depicting a dog, Blacky, going through various sexual-amative experiences with other dogs representing Blacky's father, mother and sibling. The test was based on specific Freudian hypotheses concerning an individual's psychosexual development. The test was scored using Blum's revised standard scoring system (3). To determine reliability of scoring, three psychologists were asked to score 10 Blacky protocols independently. There were only 4 disagreements about overall dimension scores among the three judges, out of a possible 130, which indicates that the scoring system employed was highly reliable.

RESULTS

The standard scoring procedure of the Blacky Pictures includes an overall dimension score for each psychosexual area which is scored either ++ (very strong), + (fairly strong) or 0 (weak or absent). The overall dimension score is derived by combining responses to the spontaneous story, the inquiry items, related comments on other cards and cartoon preferences. Due to the small number of Ss employed in this exploration, the very strong and the fairly strong overall dimension scores were combined.

Table I shows the overall dimension scores for the three ethnic groups. There were no significant differences between ethnic groups for any of the 13 psychosexual areas.

Table I reveals that 52 Ss and 47 Ss gave evidence of conflict in areas of masturbation guilt and anaclitic love object, respectively. Conversely, only 10 Ss and 14 Ss gave evidence of

TABLE I—Overall Dimension Scores for White, Negro and Puerto Rican Subjects

	Psychosexual Dimension	Weak or Absent			Very or Fairly Strong			Chi-Square*
		Wh.	Neg.	P.R.	Wh.	Neg.	P.R.	
1.	Oral Eroticism	17	11	10	3	8	8	5.48
2.	Oral Sadism	11	6	8	9	13	10	1.66
3E.	Anal Expulsiveness	10	9	5	10	10	13	3.06
3R.	Anal Retentiveness	8	13	12	12	6	6	4.10
4.	Oedipal Intensity	13	11	12	7	8	6	.45
5.	Masturbation Guilt	2	3	0	18	16	18	1.62
6.	Castration Anxiety	13	10	12	7	9	6	1.33
7.	Positive Identification	19	15	13	1	4	5	4.79
8.	Sibling Rivalry	16	14	13	4	5	5	.58
9.	Guilt Feelings	10	9	14	10	10	4	5.69
10.	Ego Ideal	7	11	6	13	8	12	3.04
11N.	Narcissistic Love Object	15	10	11	5	9	7	1.78
11A.	Anaclitic Love Object	4	3	3	16	16	15	.00

* Chi-Squares of 5.991 and 9.210 were significant at the .05 and .01 levels of probability, respectively.

TABLE II—Analysis of Mean CMAS Test Difference for All Subjects According to Presence or Absence of Each Area of Conflict

	Psychosexual Dimension	O.D.S.	CMAS Scores N	$\bar{X}$	S	t
1.	Oral Eroticism	Weak or absent	38	22.26	6.99	
		Very or fairly strong	19	25.11	7.11	2.61*
2.	Oral Sadism	Weak or absent	25	24.84	6.54	3.05*
		Very or fairly strong	32	22.03	7.29	
3E.	Anal Expulsiveness	Weak or absent	24	22.87	7.38	.63
		Very or fairly strong	33	23.45	6.90	
3R.	Anal Retentiveness	Weak or absent	33	23.18	7.89	.07
		Very or fairly strong	24	23.25	6.03	
4.	Oedipal Intensity	Weak or absent	36	22.66	7.02	1.57
		Very or fairly strong	21	24.28	7.53	
5.	Masturbation Guilt	Weak or absent	5	25.80	4.50	.54
		Very or fairly strong	52	22.96	7.32	
6.	Castration Anxiety	Weak or absent	35	23.66	7.14	1.14
		Very or fairly strong	22	22.50	7.14	
7.	Positive Identification	Weak or absent	47	23.36	7.50	.41
		Very or fairly strong	10	22.50	5.34	
8.	Sibling Rivalry	Weak or absent	43	23.09	7.08	.29
		Very or fairly strong	14	23.57	8.25	
9.	Guilt Feelings	Weak or absent	33	22.91	6.25	.78
		Very or fairly strong	24	23.63	7.62	
10.	Ego Ideal	Weak or absent	24	20.75	6.48	4.67*
		Very or fairly strong	33	25.00	7.11	
11N.	Narcissistic Love Object	Weak or absent	36	22.17	6.75	2.82*
		Very or fairly strong	21	25.00	7.50	
11A.	Anaclitic Love	Weak or absent	10	21.60	7.08	.92
		Very or fairly strong	47	23.55	7.14	

* "t's" of 2.00 and 2.66 were significant at the .05 and .01 levels of probability, respectively.

TABLE III—A Matrix of Intercorrelations between Manifest Anxiety, Lying Scores, Freedom from Psychosexual Conflict, Intelligence and Age

Number Variables	Intercorrelations 1	2	3	4	5
1. Manifest Anxiety		—.02	—.14	—.14	—.16
2. Lying Scores			—.19	—.29*	—.11
3. Freedom from Psychosexual Conflict				.12	—.27*
4. Intelligence					—.13
5. Age					

* R's of .26 and .34 were significant at the .05 and .01 levels of probability respectively.

conflict in areas of positive identification and sibling rivalry, respectively.

The range of anxiety scores was from 9 to 37 with a mean and sigma of 23.14 and 7.32. The range of L-scores was from 2 to 10 with a mean and sigma of 5.5 and 1.91. The correlation between anxiety scores and L-scores was —.02 which was not different from a zero order correlation. This latter correlation was a desirable relationship because it indicates that those Ss who falsified responses may score high as well as low on the CMAS Test.

In Table II all Ss have been categorized according to overall dimension scores on the Blacky Test. Within each category are the number of Ss and the Ss' mean and sigma scores for the CMAS Test. The data of Table II indicate four significant differences in anxiety. Ss who gave evidence of problems in the areas of oral eroticism, ego ideal, and narcissistic love object scored significantly higher in

anxiety than did Ss not giving evidence of these conflicts. In the area of oral sadism, Ss who gave no evidence of problems scored significantly higher in anxiety than Ss who gave evidence of problems. There were no significant differences in anxiety between Ss who gave evidence and those giving no evidence of problems in the nine remaining areas.

Table III shows the intercorrelations between anxiety, L-scores, intelligence, age and freedom from psychosexual conflicts. Freedom from psychosexual conflicts was estimated by counting the zeroes for the 13 psychosexual dimensions. The correlation matrix shows that anxiety was not significantly related to L-scores, freedom from conflict, intelligence nor age. Freedom from conflict, however, was significantly related to age: the older the boy, the greater the conflict. Freedom from conflict was unrelated to L-scores, intelligence and age. Lying scores were significantly related to intelligence; the more intelligent the boy, the less the lying score. Lying scores were not significantly related to age.

DISCUSSION

The main purpose of this study was to explore the general relationship between manifest anxiety and the specific areas of psychosexual conflict which were measured by the Blacky Pictures Test. This problem area has not been explored previously so that the writers had only one rationally-derived hypothesis to guide the study. This hypothesis, which was based upon psychoanalytic theory, predicted a positive relationship between manifest anxiety and specific areas of psychosexual conflict.

The main finding of this study showed *no* significant relationship between freedom from psychosexual conflicts and manifest anxiety. Only three specific areas of conflict were significantly related to manifest anxiety. Why were there no significant relationships between manifest anxiety and such theoretically related problems as guilt feelings, sibling rivalry, castration anxiety, etc.? While the answer to this question lies beyond the scope of this study, an explanation for the findings reported herein may lie in: 1) chance factors related to the distribution of the number of tests of significance used; 2) peculiarities within the delinquent sample; 3) the unreliability of the instruments; or, 4) other artifactual considerations.

While the results of this exploration probably do not possess demonstrable stability, they may serve as a source of hypotheses for further research. The following hypotheses might be proposed: 1) There may be no positive relationship between the expression of anxiety and the number of areas of psychosexual conflict; 2) the expression of anxiety may be related to specific areas of stress; and, 3) there may be a triad relationship among delinquent boys between manifest anxiety, oral eroticism and oral sadism: the greater the oral dependency and the less the conflict over the expression of anxiety, the greater the manifest anxiety.

The findings of this exploration and the discussion above should be interpreted in terms of the kinds and numbers of Ss used and the limitations of the instruments employed. Whether these findings may be generalized to dissimilar populations is a matter for future research.

SUMMARY

The purpose of this exploration was to investigate the relationship between manifest anxiety and specific areas of psychosexual conflict. Manifest anxiety was operationally defined in terms of Ss' scores on the Children's form of the Manifest Anxiety Scale. Specific areas of psychosexual conflicts were limited to Ss' performance on the 13 areas of the Blacky Pictures Test.

The Ss of the study were 57 white, Negro, and Puerto Rican boys at the

New York State Training School for Boys, a residential institution for adjudicated delinquents at Warwick, New York. The reliability correlation of these Ss on the CMAS Test was .86 which indicated that the performance of Ss was statistically reliable.

The major findings of this exploration may be enumerated as follows:

1. There were no significant differences in psychosexual problems between white, Negro and Puerto Rican Ss for any of the 13 areas studied.

2. The two most frequent areas of psychosexual problems for all the Ss were masturbation guilt and anaclitic love objects. Problems in the areas of positive identification and sibling rivalry were the most infrequent for this group.

3. Ss who scored fairly strong or very strong for the areas of oral eroticism, ego ideal and narcissistic love object were significantly more anxious than those Ss who scored weak or absent for these three dimensions.

4. In the area of oral sadism, Ss who were categorized as weak or absent were significantly more anxious than those Ss who were categorized as fairly strong or very strong for this dimension.

5. There were no significant differences in anxiety for those Ss categorized as strong or fairly strong and those Ss who were categorized as weak or absent from the following dimensions: anal expulsiveness, anal retentiveness, oedipal intensity, masturbation guilt, castration anxiety, positive identification, sibling rivalry, guilt feelings and narcissistic love object.

6. Freedom from psychosexual conflicts, which was defined in terms of a quantitative summation of those areas of absence of stress, was not significantly related to manifest anxiety.

References

1. Aronson, M. L. A study of the Freudian theory of paranoia by means of the Blacky Pictures. *J. proj. Tech.*, 1953, *17*, 3-19.

2. Blum, G. S. *A study of the psychoanalytic theory of psychosexual development.* Unpublished doctor's dissertation. Stanford University, 1949.

3. ——— A study of the psychoanalytic theory of psychosexual development. *Genet. Psychol. Monogr.*, 1949, *39*, 3-99.

4. ——— *The Blacky Pictures: a technique for the exploration of personality dynamics.* New York: Psychol. Corp., 1950.

5. Buros, O. K. (ed.). *The fourth mental measurements yearbook.* Highland Park, New Jersey: Gryphon Press, 1953.

6. Castanenda, A., McCandless, B. R., and Palermo, D. S. The children's form of the manifest anxiety scale. *Child Develpm.*, 1956, *27*, 317-332.

7. Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

8. Freud, S. *New introductory lectures on psychoanalysis.* New York: Norton, 1933.

9. —— *Inhibitions, symptoms and anxiety.* London: Hogarth Press, 1936.

10. Goodstein, L. D. Interrelationships among several measures of anxiety and hostility. *J. consult. Psychol.*, 1954, *18*, 35-39.

11. Hall, C. S. Freud's concept of anxiety. *Pastoral Psychol.*, 1955, *6*, 43-48.

12. Hoch, P. H., and Zubin, J. (Eds.) *Anxiety.* New York: Grune & Stratton, 1950.

13. Klein, D. B. *Mental hygiene.* New York: Holt, 1956.

14. Lindner, H. The Blacky Pictures Test: a study of sexual and non-sexual offenders. *J. proj. Tech.*, 1953, *17*, 79-84.

15. May, R. *The meaning of anxiety.* New York: Ronald Press, 1950.

16. McCandless, B. R., Casteneda, A., and Palermo, D. S. Anxiety in children and social status. *Child Develpm.*, In press.

17. Michal-Smith, H., et. al. Use of the Blacky Pictures with a child whose oedipal desires are close to consciousness. *J. clin. Psychol.*, 1951, *7*, 280-282.

18. Molish, H. B. Character structure of adjusted and maladjusted naval recruits as measured by the Blacky Pictures. *Amer. J. Orthopsychiat.*, 1954, *24*, 164-174.

19. Rosen, E. The Blacky Pictures by Gerald Blum. *J. proj. Tech.*, 1951, *15*, 109-110.

20. Seward, J. P. Psychoanalysis, deductive method and The Blacky Test. *J. abnorm. soc. Psychol.*, 1950, *45*, 529-535.

21. Smock, C. D., and Thompson, G. G. An inferred relationship between early childhood conflicts and anxiety responses in adult life. *J. Pers.*, 1954, *23*, 88-98.

22. Tarachow, S. Forepleasure aspects of a hobby as a defense against castration anxiety. *J. Hillside Hosp.*, 1954, *3*, 67-72.

23. Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, *48*, 285-290.

24. ——— The Taylor Manifest Anxiety Scale and intelligence. *J. abnorm. soc. Psychol.*, 1955, *51*, 347.

25. Trent, R. D. The relationship of anxiety to popularity and rejection among institutionalized delinquent boys. *Child Developm.*, In press.

Received March 18, 1957

BOOK REVIEW

Leary, Timothy. *Interpersonal Diagnosis of Personality.* New York. The Ronald Press Company 1957. $12.00.

This 518-page volume, according to the author, is concerned with interpersonal behavior primarily as it is expressed and observed in the psychotherapeutic setting. Its value "lies in its emphasis on the complexity and variety of human nature and upon the objectivity and clarity of the empirical procedures it sets forth for multi-level diagnosis". As a whole this appears to be one of the most original and stimulating documents in the area of personality theory and assessment that has appeared in some time. It focuses on what seems to this reviewer to be the central problem in clinical assessment today, that of multi-level assessment. Much of the literature on projective methods and their validity is confusing and ambiguous because the possibility that assessment may be taking place at disparate levels has been largely ignored. Thus, projective techniques which measure mainly personality areas that are below the limen of consciousness but that have no direct behavioral referrents have been found to be lacking in "validity" due to their low correlation with directly observed behavior.

Leary begins by frankly stating that his emphasis will be on the interpersonal aspects of personality. He documents the importance and significance of this aspect and adequately relates his views to currently existing theories in the field. The first major section of the book deals with the definition of the various levels and with suggested ways of measuring them. His discussion of the level of public communication (level I) is excellent and focuses on an area frequently ignored by clinical psychologists as well as psychiatrists. Many times in our usual methods of assessment we tend to assume that the private communications of the individual as seen through projective instruments, together with the conscious self-concept derived from interviews, make for a realistic and adequate picture of personality functioning. It has not been sufficiently considered that an individual may play many roles in his life and his functioning efficiency in various interpersonal situations may differ greatly from both his description of himself and his private motives, as gauged by depth procedures. The concept of "interpersonal reflex" suggested by the author seems an excellent description of the tendency that many individuals have to behave in stereotyped ways towards others and almost to force their fellow man into taking an attitude towards them based on their own conflicting needs and projections. The author's discussion of the level of conscious communication (level II) is also quite good, providing many stimulating ideas for the clinical worker. Leary emphasizes the bias implied in focusing upon hostility and aggression as signs of pathology which seems to be part of our clinical culture. He points out that "overconventionality" and "hypernormality" can also be forms of maladaptive behavior and suggests terms and concepts describing this type of interpersonal technique. When the author comes to the level of private perception (level III) he shows less strength and less sophistication. The term "preconscious" which he assigns to this level seems rather unfortunate, even though it is always stated in quotations. "Preconsciousness" has theoretical connotations which seem quite inappropriate here and tend to confuse the issue throughout the ensuing parts of the book. His assumption that the Rorschach test is useless in measuring this level and that content analysis of this test has been the "stepchild of Rorschach-type theory" seems rather gratuitous.

The other levels that the author discusses are taken up in less detail and seem to be of less significance in the interpersonal system conceived of at present. The level of the unexpressed (level IV) is discussed in a rather muddled theoretical way, and the assumption that this is a level of greater depth than the preceding level is not documented very effectively. The level of values (level V) seems to dangle in this system without any particularly convincing reason being offered for its inclusion. This section of the book concludes with a general discussion of interpersonal diagnosis which is essentially sound although somewhat defensive in tone. Why the author should be so delighted at the prospect of turning over the bulk of his work to clerks and technicians is rather difficult to understand, unless he has deluded himself into really believing that his techniques are fully adequate as a basis for making the important decisions about the lives of patients that have to be made in a psychiatric setting. His dogmatism here is

unfortunate since it detracts from the real contribution the system has to make to the total understanding of individuals, both sick and well. The suggested relationships between interpersonal diagnosis and Kraepelinian nosology are interesting but seem at this point to be mainly *a priori* and are not documented by any statistical evidence.

The next section of the book deals with what the author refers to as "the variability dimension of personality". Good operational definitions of the usual defense mechanisms are supplied in this section and are related to interpersonal indices in a plausible manner. The fact that such definitions constitute merely logical constructs of an *a priori* nature is not sufficiently emphasized, possibly leading to the false impression that they are the results of research. The clinical chapters following are entertaining and interesting, providing much demonstrative case material of value to the clinician. They seem rather less original than the preceding sections and provide little that is new and startling. Some of the correlations cited are less than impressive, since level I phonomena are often based on test categories and then in turn related to them, giving the impression of pulling oneself up by one's own bootstraps; also the overlapping and uniqueness of signs tends to be glossed over. It is rather difficult to understand where the author gets such assumptions that the MMPI F scale measures "nonconformity" or that the Ma scale measures "imperturbability".

The diagnostic sections give an impression of overemphasis on symptoms which fails to do justice to Leary's intent and may be misleading to the reader. There is some doubt in the reviewer's mind as to the value of including this material in the book. Among the more original points made in this section are (a) the emphasis on the solicitation of interpersonal behavior from others which characterizes different personality types, and (b) the influence of identifications made by people throughout their life on the development of interpersonal techniques. Leary's discussion was saved from stereotypy by the inclusion of concepts of this kind which seem highly plausible and refreshingly novel.

The last section of the book deals with some illustrative applications of the interpersonal system. It is pointed out that hospital patients show more interlevel disparity. The author fails to discuss how the use of his system is essentially different from the usual "team" approach in which the social worker may make contributions at the level of public communication, the psychiatrist at the level of conscious communication, and the clinical psychologist at the level of private perception. However, the Leary frame of reference could easily be applied to the above and would make the whole staff system of diagnosis much more theoretically comprehensible. The discussion of psychosomatic types emphasizes that different types of psychosomatically ill individuals may present the same facade but have different underlying multi-level patterns. This appears to be a real contribution to knowledge and gives some objective evidence for much clinical folklore.

The discussion of the applications of the system to industrial management also seems very helpful for diagnostic purposes. The question is left open as to how this knowledge might be used for counseling or remedial purposes. Finally, some tantalizing bits and pieces of information are given as to the application of the system to the area of group psychotherapy. It is unfortunate that no more space is devoted to the discussion of this subject since it seems to be one of the most obvious and useful ways of applying the theory and assessment procedure.

The reviewer enjoyed reading this book very much. It brings together much of the thinking and theorizing on the part of Leary and his colleagues that has been found in various articles scattered throughout the literature in the past six years. It is suggested that every clinical psychologist regardless of his level of sophistication, will find his thinking stimulated and his assessment program broadened by perusal of this volume. An incorporation of these theories might well make our research in the validation of projective techniques much more relevant and germane, permitting the definitive exploration as to what is and what is not to remain with us as a tool during the years to come.

WALTER G. KLOPFER, PH.D.
Chief Clinical Psychologist
Norfolk State Hospital
Norfolk, Nebraska

GENERAL NEWSLETTER

de Zeeuw, J. The administration and interpretation of the Rorschach test in three phases. *Zeitschr. f. diagn. Psychol.*, 1957, *5*, 5-19.

An extension of the administrative procedure is suggested for the purpose of enriching the diagnostic process with the Rorschach.

Janssen, Ingeborg. Beurteilung der Zeugentüchtigkeit eines 13 jährigen Mädchens. *Zeitschr. f. diagn. Psychol.*, 1957, *5*, 20-30.

The psychological examination of an adolescent sexual accuser suggests the falsity of her accusation and stimulates an admission.

Sazi, Morio and Kataguchi, Yasufumi. A study on the measurement of the effects of psychotherapy. *J. ment. Health*, 1956, 1-52. (Tokyo)

Ratings, therapeutic session content, and Rorschach protocols were employed in the evaluation of the nature and amount of psychotherapeutic change.

Kataguchi, Yasufumi and Dendo, Hisako. A study on the personality of the traumatic neurosis by Rorschach test. *J. ment. Health*, 1955, 30-41. (Tokyo)

Patients diagnosed as traumatic neurotics with hysterical symptoms revealed Rorschach pictures of rigid personalities rather than the conventional hysterical pictures.

Delay, Jean; Pichot, P.; Lemperiere, T.; and Perse, J. Le test de Rorschach dans le psychosyndrome organique. *Rev. de Psychol. Appliquée*, 1956, *6*, 247-287.

A summary of the research findings of various investigators of brain syndromes by means of the Rorschach. The authors find the Piotrowski signs to be more valid than the others.

ANNOUNCEMENTS

EIGHTEENTH ANNUAL MEETING SOCIETY FOR PROJECTIVE TECHNIQUES

New York 1957

PROGRAM

Thursday, August 29, 1957, 10:00 a.m.

Semi-annual meeting of the Board of Trustees, 27 W. 86th St.

Saturday, August 31, 1957

Symposia on Evaluation of the Process and Results of Therapies.

I. General Problems of Methods and Theory: Hans Strupp, Ph.D., Chairman. *(Penn Top South, Hotel Statler)*
   a. Leopold Bellak, M.D.: "Studying the Psychoanalytic Process".
   b. Helen Sargent, Ph.D.: "Studying the Outcome of Therapy".
   c. Morris Parloff, Ph.D.: "Studying the Communication of Values and Therapeutic Change".
   d. Herman Molish, Ph.D.: "Studying the Effects of Ataractic Drugs".

II. Results with Psychotherapies: Robert R. Holt, Ph.D., Chairman. *(Penn Top South, Hotel Statler.)*
   a. Gerald Ehrenreich, Ph.D.: "Assessing Therapeutic Results with Psychological Tests."
   b. Eli Rubinstein, Ph.D.: "Assessing the Results of Out-Patient Psychotherapy".
   c. Roy Schafer, Ph.D.: "Problems of Interpretation in Re-testing."

III. Results with Ataractic Drugs: *(Penn Top South, Hotel Statler).*
   a. Report of the Diagnostic Devices Committee (Division 12). Herman B. Molish, Ph.D., Joseph McGovern, Ph.D., Newell Berry, Ph.D., Benjamin Pope, Ph.D., Kenneth B. Little, Ph.D.: "A critical Review of some Research Studies concerned with the Effects of Ataractic Drugs."
   b. Samuel W. Greenhouse: "Some Methodological Aspects of the Evaluation of the Effects of Ataractic Drugs, A Review of some Controlled Studies."
   c. Herbert J. Schlesinger, Ph.D.: "The Use of Projective Techniques as Applied to the Evaluation of the Effects of the Ataractic Drugs."
   d. Albert Kurland, M.D.: "Evaluation of the Ataractic Drugs in the Psychiatric Treatment of State Hospital Patients. Some Results and Special Problems concerning Methodology."
   e. A. D. Annis, L. Mitchell, M. Zax, Ph.D.: "The Conditioned Galvanic Reflex as a Method of Evaluating the Effects of Ataractic Drugs in a Group of Psychiatric Patients."

BUSINESS MEETING, Society for Projective Techniques, *West Room, Hotel Statler,* Saturday, August 31, 1957, 4 to 4:50 p.m.

Agenda for the Annual Business Meeting

1. *Officers' Reports*
   a. Secretary ..... Helen Davidson
   b. Treasurer ..... Gordon Derner
   c. Executive Editor ..... Bertram Forer
2. *Committee Reports*
   a. Ethics ..... Margaret Mercer
   b. Finance ..... Harry McNeill
   c. International ..... Grace Cox
   d. Membership ..... Walter Klopfer
   e. Nominating ..... Jules Holzberg
   f. Program ..... Roy Schafer, Leopold Bellak
   g. Regional ..... Herbert Dorken
   h. Research ..... Herman Feifel
   i. Training ..... Walter Kass
3. *New Business*
4. *Introduction of Newly Elected Officers*

ANNUAL DINNER MEETING: Society for Projective Techniques, *Penn Top North, Hotel Statler. Price* $5.75 plus 15%, plus 3%. 6:00 p.m.

PRESIDENTIAL ADDRESS: Bruno Klopfer: "Psychological Variables in Human Cancer". About 7:45. Non-diners are also invited.

Sunday, September 1, 1957

11 a.m. to 12:50 p.m. *East Room, Hotel Statler.*

Joint Symposium with Division 12. "Research with Projective Techniques". Herman Feifel, Ph.D. Chairman.

Speakers: Leonard S. Abramson, Ph.D., Ralph H. Gundlach, Ph.D., Edith Wiesskopf-Joelson, Ph.D., Henry David, Ph.D.

Discussant: Bertram R. Forer, Ph.D.

Monday, September 2, 1957

Meeting of the newly-elected Board of Trustees at the home of Dr. Leopold Bellak, the new President, Larchmont, N. Y., 2 p.m.

For specific times and places and in the event of subsequent changes in plans, please consult your official A.P.A. program at the meetings.

REPORT BY THE INCOMING PRESIDENT

The Executive Editor is anxious to get this copy of the Journal to you before the meetings, so that you can peruse the program printed in this copy. Therefore, I am taking this early opportunity to report to you since

I particularly hope that you will attend the business meeting. This time, it will be combined with an open meeting of the Board of Trustees in order that you may have more direct contact with current work than usual. In addition, your full attendance at this meeting will be a fitting tribute to Dr. Bruno Klopfer. When he turns the presidency over to me formally, he will not only have completed service in that capacity but served as Editor of the Journal for twenty years. In these, and in other capacities he has been of outstanding value to the Society and to projective methods. He deserves our thanks.

It is my intention to inaugurate a permanent presidential page with this report since I feel that a greater contact between the membership and the officers of the Society will be useful to the development of our field. Our secretary, Dr. Helen Davidson, has familiarized you with the events of the past year; therefore, let me take this occasion to give you a preview of the plans I hope to submit to the Board.

If we define our Society's purpose as the extension of the optimal usefulness of projective techniques in the general context of psychology and allied behavioral sciences, then I believe that the following should be done immediately:

(1) *Research activities should be increased and integrated.* This can be done by establishing a "Research Registry" as part of and a function of the Research Committee. It would be a central *fact-gathering, fact-reporting* and *fact integrating* agency.

Projective records would be micro-carded; and this would be one step toward establishing normative frames of reference and facilitating comparisons and exchange of data. Without going into too great a detail: Dr. Klopfer and the Board have authorized Dr. Feifel (chairman of the Research Committee) and myself to engage in some preliminary planning with Dr. Bertram Kaplan, Secretary of the Committee on Primary Records, National Research Council, for the micro-carding of projective records.

In addition, I am trying to obtain support from other Foundations to make it possible for the Board to *stimulate* needed research and to serve in an advisory capacity to thesis supervisors and research workers.

(2) The *Training Committee* is the watchdog for the teaching of projective techniques. It is my belief that only Fellows of the Society should teach these techniques – the Society thus safeguarding the standards of proficiency. It is my aim to establish close contact between our Training Committee and the APA Committee on Certification of teaching programs.

The Training Committee, under the chairmanship of Dr. W. Kass, has made a preliminary survey of the teaching of projective techniques. Dr. T. Abel, in a private conversation, impressed me with the fact that projective techniques are probably taught too early in the clinical program and not properly integrated with the rest of the clinical experience. Dr. M. Hertz has expressed an interest in the improvement of report writing; and I firmly believe that some standards must be developed.

(4) The *Program Committee* has, this past year, put special emphasis on projective techniques as an interdisciplinary tool; and it has appeared on various conference platforms under this aegis. Efforts will be made to have again programs jointly sponsored by our Society and the American Psychiatric Association, the American Psychological Association and the Orthopsychiatric Association. The Board has already approved an ad-hoc liaison committee with other professional societies to assist them in programming work on projective techniques and the optimal use of projective techniques in the respective fields.

(5) *Finances.* It is my firm belief that the Society should make some earnest attempt to solicit grants and, particularly, bequests. The Society should have some capital, the earnings of which could be used to further the activities of projective techniques. Therefore, all members are asked to keep the Society in mind whenever they detect philanthropic urges in themselves or in others!

Above All: The creative thinking of the membership is needed. The job ahead is simple: We must improve the functioning of our Society and advance the use of projective techniques. This can only be done by full membership participation. Therefore, all suggestions will be greatly appreciated.

LEOPOLD BELLAK, M.D.

COUNCIL OF PSYCHOANALYTIC PSYCHOTHERAPY

The Council of Psychoanalytic Psychotherapy will hold its first general meeting at 1:30 p.m., October 26, 1957 at the Carnegie Endowment for International Peace, United Nations Plaza, New York City. The Council is composed of practising psychoanalysts of varied theoretical orientations. Members of this group are drawn from the fields of medicine, social work and psychology. Persons interested in joining the Council of

Psychoanalytic Psychotherapy are invited to attend the meeting. Any inquiries regarding membership should be addressed to DR. IVAN WENTWORTH-ROHR, 270 West End Ave., N.Y. 23.

REGIONAL REPORT

The New York chapter of the Society for Projective Techniques held its annual meeting and all-day conference on the subject: "Rorschach and Freud," on Saturday, May 11, 1957, at New York University. Emanuel K. Schwartz was chairman. The address was presented by Zygmunt A. Piotrowski and subsequent small group discussions were led by Fred Brown, Susan Deri, Asya L. Kadis, and Samuel B. Kutash.

More than 100 people attended, and the technique of small-group seminar discussion proved to be an exceedingly worthwhile arrangement for meetings.

The group leaders found the audience very inspired and they contributed to practical and clinical evaluation of the presentation.

Dr. Piotrowski's main idea was that the concepts and methods of Freud and Rorschach complement each other. The Rorschach is a tool without a personality theory: psychoanalysis contains a personality theory. The Rorschach, therefore, should be handled in Rorschach terms and then interpreted in psychoanalytic terms, if necessary. The analysis and description of personality can only benefit by uniting these two approaches. There are many Rorschach conclusions which have nothing to do with Freud, and are consequently, largely independent of anything said by Freud, and independent of the degree of validity of the specific Freudian statement. The combining of the methods and ideas of the two men should be done in such a way that the contributions of each stands out clearly.

Dr. Piotrowski's report of the personal life history of Rorschach made him come alive as a clinician and scientist, and as a person.

Dr. Piotrowski discussed also anxiety and emotional, personality levels, the relationship between color, form, movement, and original responses. He gave also clinical instances.

For the year 1957-58, Samuel B. Kutash was elected President of the New York Chapter, Mrs. Janet Ginandes, Secretary, and as members of the Executive Committee Theodora M. Abel and Emanuel K. Schwartz immediate past President.

RESEARCH EXCHANGE

This new column was started for the purpose of publishing speculative research ideas, of listing rare projective records which the owner is willing to lend to other psychologists, and of printing want-ads for such records. (see this journal, 1957, *21*, 103).

Bert Kaplan, from the University of Kansas, has sent us the following contribution.

Volume I of *Primary Records in Culture and Personality*, a publication on Microcards of personality materials in nonliterate and non-Western societies was published early this spring. The series is under the sponsorship of the Committee on Primary Records, National Academy of Sciences-National Research Council and is edited by Bert Kaplan, University of Kansas. Volume I which contains almost 4,000 pages of data consists of 131 Microcards of the standard 3 x 5 inch size and sells for $35.00. Volume II containing 7,600 pages will be available shortly. The series may be ordered from The Microcard Foundation, Box 2145, Madison, Wisconsin.

The first two volumes include projective test, life history and dream data collected by such workers as Victor Barnouw, Lucien M. Hanks, Jr., John Honigmann, G. Morris Carstairs, Erika Bourguignon, William Caudill, Alen Kerckhoff, Blanche Watrous, Anthony F. C. Wallace, Robert Anderson, Elizabeth Colson, Bert Kaplan, Kurt Wolff, Paul Wright, Evon Z. Vogt, Jr., Paul Hauck, John Adair, Jules Henry, Thomas W. Maretzki, Murray A. Straus, Mohamed I. Kazem, A. Irving Hallowell, Ernestine Friedl, Edward Bruner, Louise Spindler, George Spindler, Weston La Barre, Dorothy Eggan, David Horr, Alfred Johnson, Dorothea C. Leighton, William R. Holland, Fred Eggan, Hildred Geertz, Sally Lewis, Alice Joseph, Veronica Murray, Francis B. Mahoney, William A. Lessa, Melford S. Spiro and Madeline Kerr. The cultures represented in the first two volumes are Nepal, Thailand, Pakistan, Hindu and Moslem of Udaipur, India, Bhil's of India, Great Whale River Eskimo, Cree, Ojibwa, Tuscarora, Cheyenne, Pomo, Spanish-American, Hopi, Navaho, Ute, Zuni, Pilaga, Haiti, Jamaican-Chinese, Ceylon, Egypt, Mandan-Hidatsa, Menomoni, Kiowa, Mexico, Java, New Ireland, Saipan, Palau, Ulithi, Ifaluk and Jamaica.

These data will be available in more than 60 college and university libraries and constitute a rich resource for the study of cultural differences in responses to projective tests. They not only are applicable to problems of national and cultural differences in personality but are also relevant to the problem of understanding the nature of cultural factors in projective test performances.